DOCUMENT RESUME

ED 160 240 PS 010 168

AUTHOR Granville, Arthur; And Others
TITLE A Process Evaluation of Project Developmental Continuity, Interim Report VI: Recommendations for Continuing the Impact Study.
INSTITUTION High/Scope Educational Research Foundation, Ypsilanti, Mich.
SPONS AGENCY Office of Child Development (DHEW), Washington, D.C. Early Childhood Research and Evaluation Branch.
PUB DATE Mar 77
CONTRACT HEW-105-75-1114
NOTE 158p.; For related documents, see ED 144 715, PS 010 163-167, and PS 010 169-176; This series includes all the public reports generated by this study ; Parts marginally legible due to poor print quality

EDRS PRICE MF-$0.83 HC-$8.69 Plus Postage.
DESCRIPTORS Attrition (Research Studies); Comparative Analysis; Data Analysis; Data Collection; *Demonstration Programs; *Early Childhood Education; Educational Status Comparison; Evaluation Methods; Formative Evaluation; Matched Groups; *Measurement Instruments; Program Effectiveness; Test Reliability; Test Validity
IDENTIFIERS *Developmental Continuity; *Project Developmental Continuity; Project Head Start

ABSTRACT

This interim report re-examines data on instrument suitability, comparability of groups, and adequacy of sample size in Year III of the process evaluation of Project Developmental Continuity (PDC) and offers preliminary recommendations concerning the feasibility of continuing the impact study. PDC is a Head Start demonstration program aimed at providing educational and developmental continuity between children's Head Start and primary school experiences. Chapter I presents a general overview of the PDC evaluation. Chapter II describes data collection and data analysis procedures and discusses issues pertaining to validity and reliability of the evaluation measures. Chapter III presents findings in the form of tabulations of characteristics of the samples and of the evaluation instruments and assessments of sample size and attrition rate for the groups in each site. Chapter IV summarizes findings on group comparability and adequacy of the samples and instruments and considers prospects for continuing the Impact Study in the light of these findings. Appendices include descriptions of the measures in the fall battery, forms for weekly tester monitoring, commentary on scoring specific scales, and 5 additional sets of forms for data collection and analysis. (Author/CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED160240

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Project Developmental Continuity Evaluation

PS 010168

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John M. Love

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Recommendations for Continuing The Impact Study

Interim Report VI

March 1977

This report was prepared for the Early Childhood Research and Evaluation Branch, the Office of Child Development, Office of Human Development Services, Department of Health, Education and Welfare under Contract No. HEW-105-75-1114, Dr. Esther Kresh, Project Officer. Views or conclusions contained herein should not be interpreted as reflecting the official opinion of the sponsoring agency.

A PROCESS EVALUATION OF PROJECT DEVELOPMENTAL CONTINUITY, INTERIM REPORT VI: RECOMMENDATIONS FOR CONTINUING THE IMPACT STUDY

March 1977

Prepared by:

Arthur Granville

Judy McNeil
Mel Shelly
Mary Morris
Judy Meece
Sally Wacker

With the Assistance of:

Barb Bruemmer
Jodi Bruemmer
JoAnn Emmendorfer
Ann Hale
Robert Hanvey
Dorothy Kelly
Helen Kiddon
John M. Love
Pat Loy
Nancy Naylor
Jane Oden
Cathy Peterson
Leslie Ryan
Lynn Spencer
Jana von Fange

High/Scope Educational Research Foundation
600 North River Street
Ypsilanti, Michigan 48197

Table of Contents

		Page
I.	INTRODUCTION	1
	Overview of Project Developmental Continuity (PDC)	1
	Purposes of the PDC Evaluation	3
	Purposes of this Report	4
	Instrument appropriateness	5
	Comparability of groups	5
	Present sample size and projected attrition	6
	Organization of the Report	6
	Chapter II, Methods	6
	Chapter III, Findings	6
	Chapter IV, Conclusions	7
	Changes in Plans Since the Last Reporting Period	7
	Cancellation of Teacher and Parent Surveys	7
	Delayed commencement of fall testing in Florida	8
	Addition of comparison elementary schools	8
	Administration of the PPLAT in Texas	8
	Plans for the Next Reporting Period	10
	Preliminary analysis of PDC's impact on children	10
	Preliminary analysis of relationships between implementation and impact	10
	Analysis of PPLAT data	10
II.	METHODS	11
	Data Collection Procedures	11
	Field Organization	12
	Training Model	12
	Tester training	13
	Observer training	14
	Monitoring	14

Table of Contents
(continued)

Page

On-site monitoring. 14
Weekly monitoring 15
Weekly Pre-Transmittal Data Checks. 15
Recording and Scoring of Data 16
Data Collection Sequence. 16
Determining Child's Language Capabilities 16
Hindrances to Testing Schedule. 17
Data Analysis Procedures. 17
Step 1: Does the Internal Consistency Coefficient Indicate Reliability? 18
Step 2: Are the Measures Valid?. 20
Step 3: Are Reliability and Validity Constant Across Time and Samples? 22
Step 4: Does the Factor Structure Support Expectations? 23
Step 5: Are PDC and Comparison Groups Comparable?. . . 23
Step 6: Are Sample Sizes and Retention Rates Adequate? 24

III. FINDINGS. 27
Characteristics of the Samples. 27
General Description 27
Comparability of PDC and Comparison Groups. 30
Site-level findings 30
Aggregate-level findings. 30
Analysis of Attrition Trends. 32
A note on the attrition statistics reported for individual sites. 38
Characteristics of the Individually Administered Instruments 39

Table of Contents
(continued)

Page

Reliability for Cohort 2 in Fall 1976 39
Constancy of Reliability Across Time and Cohorts. . . . 39
Measures that show constant reliability 39
Measures that show increasing reliability 42
Measures that show decreasing reliability 42
Validity. 42
Characteristics Determined from Past Reporting Periods: Sensitivity to Change, Suitability for Older Children, and Relationship to Social Competence 46
Correlations with age 51
Fall-to-spring change: t tests 51
Fall-to-spring change: regression analysis 52
Summary of sensitivity to change. 52
Suitability of the instruments for use in the higher grades 52
Relationship to social competence 54
Ease of Administration. 56
Bilingual Syntax Measure (BSM). 56
Verbal Memory 56
WPPSI Block Design. 56
Draw-A-Child. 56
Verbal Fluency. 57
Preschool Interpersonal Problem-Solving Test (PIPS) . 57
Arm Coordination. 57
Pupil Observation Checklist (POCL). 57
Summary 58
Factor Structure of the Battery 58
Results for the English-dominant sample 58
Results for the Spanish-dominant sample 58
Comparisons with previous factor analyses 59

Table of Contents
(continued)

Page

Characteristics of the Classroom Observation System . . 61
Summary of Instrument Development 62
Purpose of this Analysis. 63
Observation Procedures. 63
Fall 1976 Observation Training Procedures 64
Reliability of the Observation System 65
Collection of reliability data. 65
Analysis of reliability data. 65
Reliability results 66
Analysis of Classroom Observation Data. 66
Preparation of observation data for descriptive analysis. 66
Results of descriptive analyses 68
Results of correlational analyses 76
Results of comparability analyses 79

IV. CONCLUSIONS 81
Adequacy of the Tests and the Samples 81
Are the Measuring Instruments Appropriate to the Task?. 81
Child measures, individually. 81
Child measures, collectively. 81
The PDC Classroom Observation System. 83
Other measures. 83
Are the PDC and Comparison Groups Really Comparable?. . 83
Overall findings. 83
Prospects for analysis at the aggregate level 84
Prospects for analysis at the site level. 84
Will Large Enough Samples of Children Remain in PDC and Comparison Schools at each Site to Permit a Longitudinal Study of Program Effects?. 85

Table of Contents
(continued)

Page

Prospects for Testing OCD's Hypotheses Regarding PDC's Impact on Children. 85

APPENDIX A: DESCRIPTIONS OF THE MEASURES IN THE FALL BATTERY 89

APPENDIX B: FORMS FOR WEEKLY TESTER MONITORING 95

APPENDIX C: COMMENTARY ON SCORING THE MCCARTHY ARM COORDINATION SCALE 105

APPENDIX D: FLOW CHARTS FOR THE ANALYSIS PROCEDURE . . .109

APPENDIX E: MAGNITUDE OF DIFFERENCES FOR VARIABLES ON WHICH GROUPS WERE FOUND UNEQUAL, BY SITE . .117

APPENDIX F: PDC CLASSROOM OBSERVATION SYSTEM: DEFINITIONS OF CATEGORIES AND MEANS AND STANDARD DEVIATIONS OF VARIABLES. 135

APPENDIX G: SUBSCALES OF THE POCL. 147

APPENDIX H: ANALYTIC PROCEDURE FOR INVESTIGATION OF PDC AND COMPARISON CLASSROOM COMPARABILITY ON THE OBSERVATION VARIABLES. 151

List of Tables

		Page
Table 1:	Composition of the Samples	28
Table 2:	Types and Frequencies of Handicaps Reported Among Children in the Full Testing Sample.	29
Table 3:	Comparability of PDC and Comparison Groups at Each Site and of Groups Aggregated Across Sites.	31
Table 4:	Projected Retention of Cohort 2 Children for Each Year of the Prospective Longitudinal Study.	37
Table 5:	Reliability of the Child Measures: Cronbach's Alpha (Internal Consistency) for Fall 1976 Head Start Children	40
Table 6:	Comparisons of Reliability (Internal Consistency) for Cohort 1 (Tested Fall 1975 and Spring 1976) and Cohort 2 (Tested Fall 1976).	41
Table 7:	Hypothesized Correlation Matrix for PDC Child Measures	43
Table 8:	Intercorrelations of Child Measures for English-Dominant Sample: PDC Fall Data, 1976.	44
Table 9:	Intercorrelations of Child Measures for Spanish-Dominant Sample: PDC Fall Data, 1976.	45
Table 10:	Deviations of Child Measure Correlations from Hypothesized Correlations, English-Dominant Sample	47
Table 11:	Deviations of Child Measure Correlations from Hypothesized Correlations, Spanish-Dominant Sample	48
Table 12:	Factor Analysis of Scores on Child Measures, English-Dominant Head Start Children	59
Table 13:	Factor Analysis of Scores on Child Measures, Spanish-Dominant Head Start Children	60
Table 14:	Coding Reliability	67
Table 15:	Intercorrelations of Observation Variables	77
Table 16:	Correlation of Observation Variables and Other Reliable Measures, Spring 1976	78
Table 17:	Summary of Findings on Characteristics of Tests Included in the Fall 1976 Battery.	82

List of Figures

		Page
Figure 1:	Master Flow Chart for Analysis and Interpretation of Data.	19
Figure 2:	Estimated Retention/Attrition Curve, Head Start through Grade 3, for Head Start Populations at PDC Sites, Based on Data for Groups that Preceded Cohorts 1 and 2	25
Figure 3:	Background Characteristics of Aggregated Analytic Samples of English-Dominant PDC and Comparison Children	33
Figure 4:	Background Characteristics of Aggregated Analytic Samples of Spanish-Dominant PDC and Comparison Children	34
Figure 5:	Mean Standard Scores of Aggregated Analytic Samples of English-Dominant PDC and Comparison Children on Selected Demographic and Performance Measures.	35
Figure 6:	Mean Standard Scores of Aggregated Analytic Samples of Spanish-Dominant PDC and Comparison Children on Selected Demographic and Performance Measures.	36
Figure 7:	Validity Profiles for Child Measures, for two Head Start Cohorts at a Total of Three Time Points.	49
Figure 8:	Relative Frequencies of Classroom Involvement by Classroom Activity Level	69
Figure 9:	Classroom Verbal Behavior: English-Speaking Classrooms.	70
Figure 10:	Classroom Verbal Behavior: Spanish-Speaking Classrooms.	71
Figure 11:	Relative Frequencies of Classroom Child-Peer Interactions: Nature of Interaction by Classroom Activity Level.	73
Figure 12:	Relative Frequencies of Classroom Child-Adult Interactions: Nature of Interaction by Classroom Activity Level.	74
Figure 13:	Relative Frequencies of Classroom Adult and Peer Interaction: Purpose of Interaction by Classroom Activity Level.	75

I

INTRODUCTION

Overview of Project Developmental Continuity (PDC)

The Office of Child Development originated Project Developmental Continuity (PDC) in 1974 as a Head Start demonstration program "aimed at promoting greater continuity of education and comprehensive child development services for children as they make the transition from preschool to school." The single most important effect of this undertaking, it is hoped, will be to enhance the social competence of the children served--that is, to increase their everyday effectiveness in dealing with their environment (at school, at home, in the community, and in society). PDC also aims to bring about broader and more intensive involvement of parents and teachers in the governance of school affairs, and to promote positive change in the institutional process, even beyond the people who may occupy the institution at a given time.

As part of the overall Head Start improvement and innovation effort, PDC emphasizes the involvement of administrators, classroom staff, and parents in formulating educational goals and developing a comprehensive curriculum. The object is to ensure that children receive continuous individualized attention as they progress from Head Start through the early primary grades. If the program is successful, existing discontinuities between Head Start and elementary school experiences will be reduced by PDC mechanisms that encourage communication and mutual decision-making among preschool and elementary school teachers, administrators, and parents.

Two program models provide alternative ways of establishing the administrative structure for continuity. In the Preschool-School Linkages approach, administratively separate Head Start and elementary school programs are brought together by the device of a PDC Council, whose membership includes teachers, parents, and administrators from both organizations. In the Early Childhood Schools approach, Head Start and elementary

school programs are combined both administratively (by the Council) and physically (in the same building), creating a new institution. In both approaches a qualitatively different program is expected to emerge as a result of Head Start-elementary school cooperation.

Continuity is expected to be established in two contexts: that of the individual child and that of the school structure. In the first context, continuity means, for example, that a child should not have to have his or her personal nature and needs rediscovered each year as he or she moves from one grade to the next; instead the child should become a more fully recognized member of the school "family" as time passes. In the context of school structure, continuity implies a cooperative pursuit of common goals, and this involves articulation of philosophies and methods in all the various areas of school enterprise. It is expected that structural continuity will contribute directly to continuity in the attention given to individual children.

School organizations at 15 sites around the country received OCD funding during 1974-75 (Program Year I) to design and plan future implementation of seven prescribed components of PDC:

- Administration: administrative coordination between and within Head Start and elementary school;
- Education: coordination of curriculum approaches and educational goals;
- Training: preservice and inservice training for teachers and childrearing training for parents;
- Developmental Support Services: comprehensive services (medical, nutritional, and social) to children and families;
- Parent Involvement: parent participation in policy-making, home-school activities, and classroom visits or volunteering;
- Services for the Handicapped: services for handicapped children and children with learning disabilities;
- Bilingual/Bicultural and Multicultural Education: programs for bilingual/bicultural or multicultural children.

During Year II, 1975-76, 14 sites (one had withdrawn voluntarily), comprising a total of 42 Head Start centers and elementary schools, began their "start-up" year, pilot-testing their adaptations of the PDC program according to plans they had drawn up during Program Year I. The program is currently in Year III, 1976-77, and PDC is now supposed to exist in mature form at the 13 participating sites (a second site withdrew at the end of Year II). If a longitudinal study of PDC is commissioned, the program's effects will be examined throughout the period beginning with the present year and continuing until the end of the 1980-81 school year. This is the period during which children in the current testing samples (Cohort 2) will progress from Head Start through grade 3.

Purposes of the PDC Evaluation

The major purpose of the PDC evaluation is to aid the Office of Child Development in its efforts to design effective programs for early childhood education. To accomplish this, the evaluation will ultimately have to provide answers to the following critical questions about PDC's impact:

- How does PDC affect children's social competence?
- How does PDC affect parents?
- How does PDC affect the attitudes and workstyles of teachers and other staff?
- How does PDC affect the school organization in terms of philosophy, methods, and social climate?

In addition to describing the consequences of PDC, the evaluation will describe and analyze the processes that led to those consequences. Although the assessment of children's social competence is very important and is emphasized in the present report, the relationship of this to the rest of the evaluation should not be neglected. Volume 2 of Interim Report IV (August 1976) delineates the process evaluation more fully; it is sufficient to emphasize here that the aims of the total evaluation are to produce conclusions about what happened (impact) and how and why it happened (process). This information will facilitate future decisions about whether the program should be replicated, and if so, how replication can best be accomplished in the light of past experience.

Purposes of this Report

Last year (Program Year II), a sample of Head Start children participated in pilot-testing of the instruments tentatively selected for the PDC battery. Demographic data were also collected for these children, and information regarding grade-to-grade attrition for children in past years was gathered from each site. This information provided preliminary answers to three critical questions:

1. Are the measuring instruments appropriate to the task?

2. Are the PDC and comparison groups really comparable?

3. Will large enough samples of children remain in PDC and comparison schools at each site to permit a longitudinal study of program effects?

On the basis of answers gained last year to question 1, some instruments were eliminated from the battery and others were modified. On the basis of last year's findings with regard to question 2, recommendations were made to sites concerning procedures that could be followed during the Head Start enrollment period to balance PDC and comparison groups in certain important respects. And on the basis of the projections made in connection with question 3, the likelihood was assessed that the samples would remain large enough over a five-year term to permit continuing statistical analysis of their performance.[1]

These same issues--instrument suitability, comparability of groups, and adequacy of sample size--are examined again in this report using data gathered in the fall of Year III. The purpose of this re-examination is to verify preliminary indications from Year II. These findings and others are integrated in this report, and their implications for a longitudinal study of PDC's impact are considered. The discussion concludes with preliminary recommendations for the future of the evaluation.

Following are brief statements of the rationales for addressing the particular issues dealt with in this report.

---

[1] These conclusions are documented and discussed in Interim Report IV, Volume 1, Pilot Year Impact Study: Instrument Characteristics and Attrition Trends, August 1976.

Instrument appropriateness. Since the ultimate goal of PDC is to enhance the social competence of children, it is essential to the evaluation that the instruments used yield measures that, collectively, represent social competence in an accurate and meaningful way. The criteria used in this and past reports for judging the adequacy of the instruments include:

- internal consistency (the extent to which an instrument's constituent items agree on what they reflect about an individual,
- stability (correspondence between total scores on the same instrument given to the same children at two time-points),
- validity (the degree to which an instrument measures what it is supposed to measure),
- sensitivity to change (an instrument's responsiveness to change over time in the characteristics measured),
- relevance to social competence (the degree to which the information produced by an instrument contributes to knowledge of a child's social competence),
- developmental span (suitability for children across the age range prospectively covered by the PDC evaluation), and
- ease of administration (the cost in time, effort, and resources of administering an instrument).

Comparability of groups. The effects of PDC upon children will be determined primarily by comparing the performance of children in PDC testing samples with the performance of children who are similar, but who do not participate in PDC (i.e., a comparison group). The assumption implicit in this is that the children in the two groups would remain parallel were it not for the intervention of PDC, and thus the way children in the comparison group perform in the future stands for the way PDC children would have performed without the presumed advantage of PDC. Whether this assumption itself stands or falls depends upon the initial equivalence of the two groups. Unless they are very similar to begin

with, or can legitimately be equalized by statistical means, no sensible comparison can be made. The similarity of PDC and comparison groups at each site and at all sites collectively is examined in this report as a check on the premise of group comparability.

Present sample size and projected attrition. In addition to the requirement of comparability, it is also important that the PDC and comparison testing samples remain large enough as time passes to permit continuing analyses of their relative performance. Attrition will occur inevitably, and the smaller the groups become, the more difficult it will be to separate PDC's effects from the effects of the many other factors that contribute to the performance of the children involved.

In the fall of this year, Program Year III, a check was made on the number of children from the Year II PDC and comparison Head Start testing samples who had enrolled in kindergarten at PDC and comparison elementary schools, respectively. Non-enrollment, obviously, constitutes attrition. Since last year's children (Cohort 1) come from the same populations as this year's target children (Cohort 2), the Cohort 1 attrition rate provides a basis for estimating the Cohort 2 rate. Projections of year-to-year attrition from Cohort 2 are presented and discussed in this report.

Organization of the Report

Chapter II, Methods. The data collection and data analysis procedures followed during the current phase of the Impact Study are described here. These descriptions document the origins of the data presented in this report.

Chapter III, Findings. This chapter presents:

- tabulations of critical characteristics of the instruments in the battery,
- comparisons of the characteristics of the PDC and comparison groups, both by site and collectively,
- assessments of sample size and attrition rate for the groups in each site.

Chapter IV, Conclusions. Findings on group comparability, adequacy of the samples, and adequacy of the instruments are summarized here, and prospects for continuation of the Impact Study are considered in the light of these findings. The crux of the deliberations is whether or not it will be possible to provide satisfactory tests of OCD's major hypotheses about PDC's effects on children.

Changes in Plans Since the Last Reporting Period

Cancellation of Teacher and Parent Surveys. As has been noted, it is expected that children, parents, teachers, and the school organization will all be affected positively by PDC, and the evaluation has been designed expressly to provide for assessment of the program's consequences in each of these four domains. Large-scale surveys of teachers and parents were to have been the principal means of measuring the impact felt in these groups. The surveys originally were to have been conducted in spring 1976. All the necessary preparations were made, but plans were postponed because of delays in the process of clearing the forms with the Office of Management and Budget (OMB). The surveys were rescheduled for spring 1977, and although OMB approval had still not been received by the first of this year, preparations were made once again, in anticipation of approval. However, in February OMB announced that the survey plans would not be approved at all, nor would plans be approved for a full-scale study of program implementation in PDC and comparison schools. The reasons for the decision have not yet been stated, thus no account of them can be offered here.

Although it is possible that the decision, or some part of it, will be reversed, or that alternative plans might be approved, it is unlikely that sufficient time will remain in this year to conduct the surveys. The next Impact Study report was to have contained an analysis of responses to these surveys, presented in terms of differences between PDC versus comparison teachers and PDC versus comparison parents, and in terms of connections discovered between program features and teacher/parent attitudes. Now, as critical as this information may be to a decision about PDC's future, it will be inaccessible, at least for the present year.

Delayed commencement of fall testing in Florida. Last year, concerns arose about the size of the Florida PDC and comparison samples, and about attrition among the children there. These concerns led to deliberation over the site's status in the evaluation. To resolve the issue, OCD and High/Scope conducted a joint study on-site in October 1976. The conclusion of the study was that the number of children likely to remain enrolled from Head Start through grade 3 is at least as high in Florida as it is in other sites. Thus it was decided that Florida would remain in the evaluation. However, this decision was made about one month after the time when preparations for fall testing ordinarily would have begun. Consequently, testing began three weeks later in Florida than it did elsewhere, and the data could not be processed in time for this report. An account of the results of testing there will be presented in the August 1977 Impact Study report.

Addition of comparison elementary schools. Since the last reporting period, some changes have been made in designation of comparison elementary schools for three sites. These changes were initiated jointly by the PDC coordinator at each site and High/Scope to ensure that when children in the comparison Head Start testing samples progress to elementary school, the greatest number possible will go to participating comparison schools rather than to non-participating schools. (Children who go to designated comparison schools can be tested periodically over the longitudinal term of the study, but children who go to other schools may be lost to the study.) The following schools have thus been added to the lists that appeared in Appendix F of Interim Report IV, Volume 1:

Site	Added Comparison Schools
Iowa	Cassidy, Edmond, Phillips, Stowe
Michigan	Malkim, Rogers
Utah	Liberty, Lowell

Administration of the PPLAT in Texas. In spring 1976, the Preschool Productive Language Assessment Task (PPLAT), an experimental procedure developed at High/Scope, was administered to a small number of the children in the Connecticut and Iowa testing samples. (The PPLAT is designed

to yield measures that characterize children's spontaneous use of language in social situations.) The resulting data were analyzed in conjunction with other PDC data that had already been obtained for the same children. The purpose of this analysis was to explore the merits of incorporating the PPLAT in the PDC battery. The investigation produced the conclusion that the procedure does yield measures of language behavior that complement the other PDC measures in interesting and potentially useful ways. For example, certain PPLAT variables seem to be related to teachers' ratings of children's social demeanor, but not at all to the children's performance on cognitive tests, while other variables produced by the same PPLAT procedure seem to be related to test performance but not to teachers' ratings. Although the sample was too small and the analytic procedure too exploratory to warrant formal presentation of results here, the PPLAT was found sufficiently promising to justify its continued refinement as a tool for the PDC evaluation. Accordingly, it was decided to continue administration of the PPLAT (with some modifications, made in light of the recent findings) at a single site.

The decision to continue the PPLAT at one site rather than all, or rather than none, was based upon a concern for using the evaluation's resources to best advantage. Since the test's validity is yet to be demonstrated, it would be premature to administer it to all the children who receive the other tests, but because the PDC battery lacks a naturalistic measure of language proficiency and because this one is promising, there is much to learn from its continued application and refinement.

Since development of language proficiency receives greatest emphasis at the PDC Bilingual/Bicultural Demonstration sites, it was deemed most practical to continue the use of the PPLAT in Texas, the site with the largest number of Spanish-speaking (and thus potentially bilingual) students. Administration of the PPLAT was begun in early February at that site and will be completed by early March 1977. The procedure is conducted in each child's dominant language, either Spanish or English (as indicated by earlier PDC assessments), and also in the alternate language, if the child possesses that capability.

Plans for the Next Reporting Period

Preliminary analysis of PDC's impact on children. By spring 1977, children in the PDC Head Start testing samples will have participated for nearly a full year in the PDC program, and it will be possible then to conduct meaningful tests of the hypothesis of PDC group gain over the comparison group. The next Impact Study report, scheduled for August 1977, will present the results of this first test of the hypothesis.

Preliminary analysis of relationships between implementation and impact. The August report will also present the results of an analytic search for relationships between implementation variables and impact variables. This analysis will seek to establish connections between measures of desired outcomes (primarily, children's test performance) and the degree to which various aspects of the program have been realized from site to site (as determined by the Implementation Study).

Analysis of PPLAT data. The next report will include an account of the first formal analysis of data from the PPLAT administered in Spanish or English or both to all children in the PDC and comparison Head Start samples in Texas.

II

METHODS[1]

Data Collection Procedures

To establish a data collection routine that would yield data of the highest possible quality, the following procedures were instituted:

1. An organizational structure for individuals involved in the data collection effort was outlined, role responsibilities were defined, and a detailed training manual was produced.

2. A training model was designed that specified tester performance standards and provided for a 6- to 8-day tester training session with large-group, small-group and individualized instruction, daily reviews of each tester's performance, and discussions of potential problems.

3. On-site monitoring of testers by trainers was conducted prior to the start of the actual testing.

4. During the data collection period, testers were responsible for monitoring each other's performance on a weekly basis.

5. Site coordinators collected completed data each week and checked it for obvious errors or omissions before sending it to the High/Scope Foundation.

Each of these procedures is discussed below.

---

[1]The tests and other instruments used, and the order of their administration, are described in Appendix A. Further details on testing, monitoring, and other procedures followed by testers on site can be found in the Field Procedures Manual (High/Scope, September 1976).

Field Organization[1]

The roles of the personnel who conducted field data collection were explicitly defined in the Field Procedures Manual in order to systematize responsibilities. For example, site coordinator responsibilities included contacting the PDC coordinator regarding the start of testing; setting up and chairing a meeting with the Head Start teachers involved in the evaluation, or contacting them individually; keeping in contact with the supervisor of field operations about the status of data collection and any problems that the site was having; checking all completed data on a weekly basis; keeping up-to-date records on the status of the data collection; carrying out any needed training; testing (and, in some cases, observing) children; and monitoring testers.

Training Model

Two training sessions were held in early September 1976--a 2½-day session for the eight tester trainers and a 6-day (or, in some cases, 8-day) session for 34 testers/observers the following week. Since all but two of the eight trainers were experienced, the 2½-day trainer training session was more of a "refresher" course, consisting of a review of the child measures and practice in test administration. Because the decision to include the Florida PDC site in the fall 1976 data collection effort was not made until October, training for the two Florida testers was conducted on-site in mid-October and the observation data were collected by a High/Scope staff member at the end of the month.

---

[1]The fall data collection started the week of September 12 at all sites except Florida. The length of the data collection period was fairly constant across sites, with most testers finishing within nine or ten weeks. In Florida, testing began in mid-October and took 10 weeks to complete. Of the 36 testers involved in the fall 1976 testing, 26 were experienced PDC testers and 10 were newly hired. In two sites, Texas and California, all the testers were new while in many of the other sites an additional person was hired because of the anticipated increase in sample size. The number of testers per site ranged from 2 to 4 depending on the sample size.

Tester training. During the 8-day tester training session, each test was explained and demonstrated to the entire group; this was followed by small-group practice. The initial small-group practice session involved the use of test "scripts." The scripts consisted of test instructions, child responses, and rationales for scoring. In using the scripts, two testers would pair up and one, the "child," would perform as indicated on the script while the other tester administered the test without the script. This provided an excellent learning situation since the child responses included on the script covered all the administration rules and gave the testers a chance to work with and correct each other without having to have a trainer nearby to answer all their questions. Two scripts were written for most of the tests. Whenever possible, new testers were paired with returning testers so that they could learn from them.

Since it is critical that testers administer the tests in a standard way, each tester was systematically "checked out" on all of the child measures throughout the training session. During this procedure a trainer played the role of the child (also recording the "child's" responses) while a tester administered one or more of the child measures to her. Prior to these check-outs the trainers had decided how a trainer (acting as the child) would respond to each item on each test. This was done for two reasons: 1) to insure that each tester was exposed to the same situations, and 2) to incorporate child responses that covered all test administration directions. For example, on the PIPS interview, there are specific things for the tester to say if the child gives an unrelated answer, a repeated answer, refuses to answer, and so on. By exhibiting all these behaviors in the check-out situation, trainers were able to assess the tester's understanding and expertise in administering each of the child measures.

Standards were set for acceptable performance during the tester check-outs, and if these standards could not be met, additional training and practice was prescribed. Check-outs were then repeated at a later time during the training session to insure correct test administration.

The eight tester trainers met every evening during the training session to discuss the day's activities and to report on the progress or status of each tester. Potential problems were identified and discussed during these meetings, as were necessary schedule changes.

Observer training. The observer training session was held two days prior to the tester training session. Thus, the tester/observers attended an 8-day session instead of the regular 6-day training session. Because it would be very difficult for each tester to master both the observation system and the child tests in the amount of time allotted for training, the decision was made to have only one person from each site collect the observation data in addition to administering child tests. In all but two of the sites, an experienced tester (one who had collected observation data in the past) was selected to make the classroom observations. The new testers in Texas and California, however, had to master both an unfamiliar observation system and an equally unfamiliar test battery.

The revised classroom observation instrument was introduced to the observers as a group, with trainers demonstrating the behaviors reflected on the instrument and the observers identifying the behaviors. After the PDC Observation Training Manual was reviewed and discussed, the observers, in groups of twos and threes, wrote scripts incorporating various behaviors from the observation system categories. They then acted out the scripts while the other observers coded the behaviors. A better understanding of the behavioral categories and items was gained by this script-writing and acting. Some videotape coding of preschool-aged children in classroom settings was also done, primarily to familiarize the observers with the videotape system and the audio "beep" which signals observation intervals prior to their coding of the scenes on the reliability tape. The reliability tape, which consisted of trainers role-playing classroom scenes, was coded by all observers mid-way through the training session.

Monitoring

On-site monitoring. At all sites, data collection activities started the week following the training session. First, the testers met with the PDC coordinator and the Head Start teachers involved and spent enough time in the classrooms to become acquainted with the children. Also, new testers practiced administering the tests to one or more children (children from outside the PDC and comparison samples) during this time.

During the week of September 20 a trainer visited each site for two days to monitor individual testers as they actually worked with a child. Although it was not feasible for the testers to administer the tests to children at their Michigan training session, it was important that each tester have such experience and the opportunity to receive feedback before starting the data collection. After watching a tester, the trainer provided any additional feedback to the tester that was necessary on improving his/her interactions with the children. This procedure served two purposes: it gave the trainer an indication of how well individual testers could establish rapport and interact with children, and it helped alleviate some of the anxieties the inexperienced testers felt about administering the measures to children. No testing of PDC or comparison children was done prior to this monitoring visit.

Weekly monitoring. During the course of a testing week, testers alternately monitored each other; the one acting as monitor simultaneously completed the test booklets and the individual monitoring forms for each test. After the session, the monitor and tester discussed any errors, and the monitoring booklets and forms were sent to the supervisor of field operations at the High/Scope Foundation to be reviewed. The monitoring forms can be found in Appendix B. The categories on the forms beside which an X appears are those in which testers, as a group, made more errors than expected or than was judged tolerable. These areas are discussed in more detail in a later section, Ease of Administration.

Weekly Pre-Transmittal Data Checks

Testers were required to give or send their completed data to the site coordinator at the end of each week. The site coordinator then checked these tests, plus any she/he had completed, for recording/scoring errors. (Site coordinators and interviewers had reviewed a checklist specifying what to look for when reviewing each completed booklet, e.g., "Is the identification complete?" "Did the interviewer fail to give a second trial when it should have been given?" "Did the interviewer skip an item?") Errors were pointed out to the particular tester and, if necessary, further training was provided by the site coordinator. The site coordinator also kept track of all completed data (in addition to the individual records each tester kept of his/her classes) and mailed the completed data to the High/Scope Foundation on a weekly basis.

Recording and Scoring of Data

In addition to the site coordinators' pre-submittal check, most data collected by inexperienced testers were also checked by the supervisor of field operations at the High/Scope Foundation. Errors in recording or coding were identified and explained to the site coordinator, who then discussed them with the other testers.

Once the raw data had been screened for accuracy at High/Scope, they were sent to the data processing section to be tagged with unique identification numbers for each student, scored and verified, then keypunched and verified.

Data Collection Sequence

To facilitate data collection, each PDC and comparison class was assigned to one or more of the site's testers (depending on the number of classes at a particular site) and the order of classes to be completed was specified. Two factors were taken into account in making these class assignments: 1) the data had to be collected simultaneously in the PDC and comparison schools to insure that both groups were observed/tested during the same time period, and 2) each tester had to be assigned to test both PDC and comparison children, thus eliminating the possibility of tester bias for either group.

Determining Child's Language Capabilities

The procedure followed by testers in determining the language capabilities of children in the testing sample was 1) to ask the child's classroom teacher for his or her judgment, 2) to observe the child's verbal behavior in natural classroom conditions, and 3) on the basis of these indications, to administer the English or Spanish version, or both, of the Bilingual Syntax Measure (BSM). In most cases, this screening process produced consistent conclusions, and subsequent testing was accordingly conducted in English or Spanish or both. (In some cases this screening process led to the conclusion that a child was proficient in some third language, but not English or Spanish. These children were excluded from the testing sample.) When the screening process proved inconclusive, the tester carefully weighed all available information to reach a conclusion about the child's language capabilities.

Hindrances to Testing Schedule

As was the case during last year's data collection effort, there were some minor setbacks in getting the data collection underway, such as lack of permission slips or testing space. Also, since Head Start enrollment had not been concluded by late September, the testers themselves had to make their own rosters and add children to them as they were enrolled. And in some sites the PDC program had not identified which comparison Head Start children would be attending comparison elementary schools, and testers had to spend time looking up addresses and subsequent elementary school designations so that only those children who would be going to the designated comparison elementary schools would be tested.

Data Analysis Procedures

Interim Reports III and IV (March and August 1976) have previously presented analyses of the adequacy of the instrument battery for use in a longitudinal evaluation of PDC. These analyses were based primarily upon data obtained for a pilot sample of children (Cohort 1) who preceded this year's target sample (Cohort 2) in PDC and comparison Head Start centers. Since it is Cohort 2 that actually will be followed in the prospective longitudinal study, the analyses described in this report were performed to verify that the earlier results hold for the present sample.

The analyses were conducted in two stages, the first of which focused on characteristics of the instruments, the second on characteristics of the samples. Proceeding in sequential steps from the beginning of Stage 1 to the end of Stage 2, the analyses examined:

- reliability of the instruments,
- validity of the instruments,
- cross-time and cross-sample congruence of reliability and validity findings (examination of fall 1976 data in relation to data from fall 1975 and spring 1976),
- factor structure of the battery,

- comparability of PDC and comparison samples, and
- adequacy of present sample sizes in view of projected attrition.

These steps, outlined in Figure 1, are described in greater detail in the following text. Flow charts illustrating each of these descriptions appear in Appendix D.[1]

Step 1: Does the Internal Consistency Coefficient Indicate Reliability?

Cronbach's alpha[2], a measure of internal consistency, is an index of the amount of overlapping variance among the items that comprise a scale; it may be conceived as the mean of all possible split-half correlations (e.g., a correlation of the odd and even items within a scale). Coefficient alpha is a measure of reliability in the sense that it reflects the degree to which the constituent items of a test relate to a unitary construct--the degree to which they tend to measure the same thing.

---

[1]The psychometric analysis procedures described in this section refer to all the instruments in the battery except the Preschool Interpersonal Problem-Solving Test (PIPS) and the Classroom Observation System. Internal consistency analysis is inappropriate for the PIPS, so Steps 1 and 2 were omitted from the PIPS analysis sequence. The analysis of the Classroom Observation System is treated separately in Chapter III. Psychometric analyses were of course preceded by scoring of the tests, and a commentary on the scoring of some of the tests is presented in Appendix C.

[2]Cronbach's alpha is a generalized form of the Kuder-Richardson formula 20. The generalized formula is:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\Sigma V_i}{V_t}\right)$$

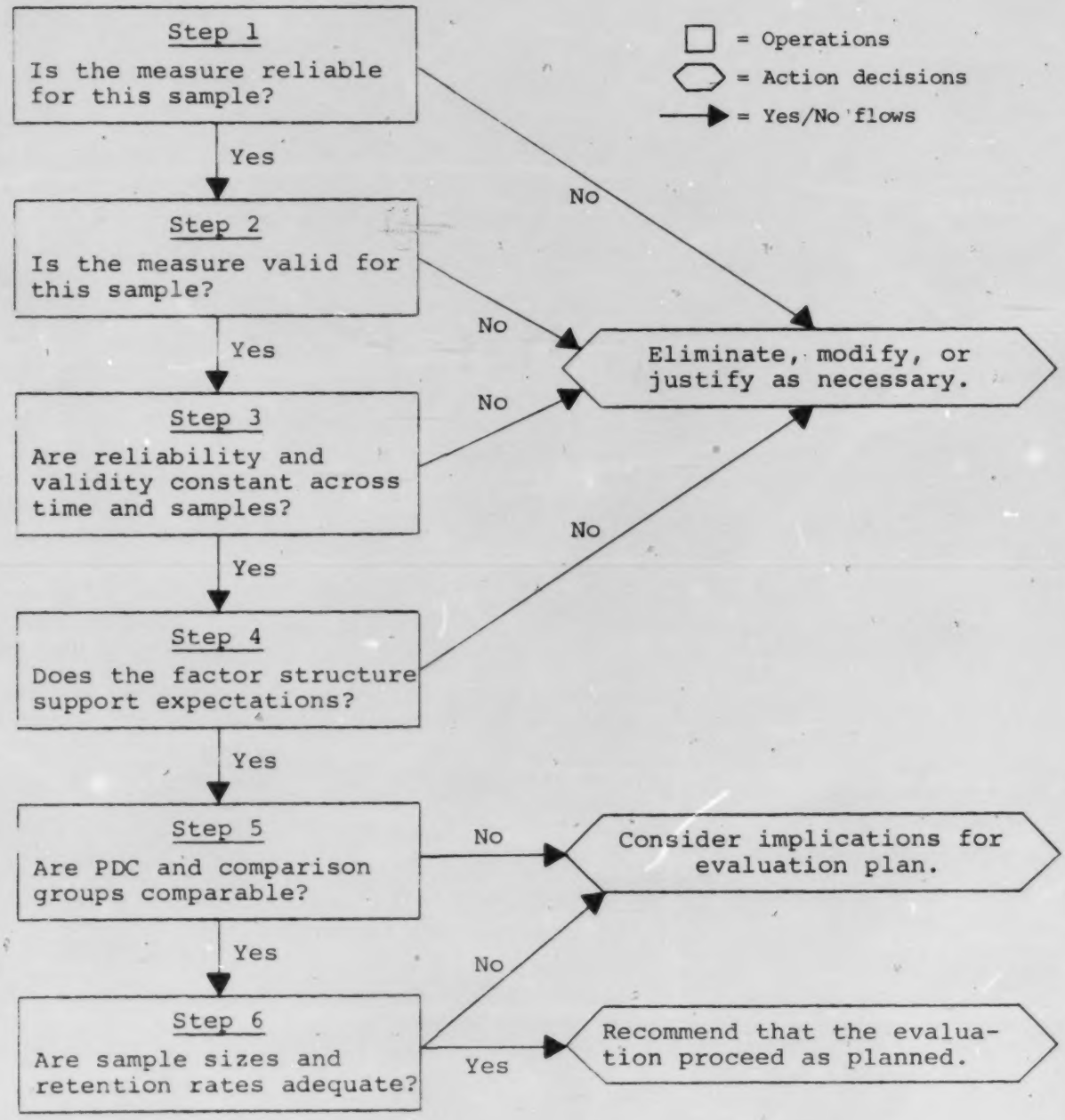
where:

n = number of items
V_i = item variance (after *a priori* weighting)
V_t = variance of the scale

Cf. Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Figure 1

Master Flow Chart for Analysis and Interpretation of Data



The procedure followed for the determination of fall 1976 internal consistency is pictured in Figure D-1, Appendix D: internal consistency was determined for each measure for the aggregated samples of English-dominant and Spanish-dominant children taking each test.[1] If the alpha for a measure was greater than .65 within either language sample, the measure was considered internally consistent for that group; if the alpha value for a measure fell below .65, that measure was likely to be dropped from the analysis procedure.

Cronbach considers alpha to be a coefficient of equivalence, which, for a test composed of relatively homogeneous items, approximates parallel-form reliability. This should be distinguished from a coefficient of stability, such as a test-retest correlation over a time interval. Cronbach himself does not set a criterion level for alpha (.65 has been selected for this evaluation, based upon a review of norms).

Cronbach suggests that a test whose items measure more than one dimension (a heterogeneous mix of items), having a low common-factor variance and therefore a low alpha, might still have a substantial test-retest correlation, and be considered stable, reliable, and interpretable. In other words, alpha is not the ultimate form of reliability. It is, rather, a useful index for estimating reliability of a homogeneous test at a single administration.

Step 2: Are the Measures Valid?

The procedures followed for determining validity are pictured in Figure D-2, Appendix D. As with reliability, prior research provides some information on the validity of the measures, but validity also must be ascertained within the context of the PDC evaluation. Most measures were selected from larger existing batteries, and items on

---

[1]The English-dominant sample includes all English-dominant, non-handicapped students for whom each measure was obtained. The Spanish-dominant sample includes all non-handicapped, Spanish-dominant students for whom each measure was obtained in Texas and California. The procedure for determining language dominance is described above under Data Collection Procedures.

most of the measures have been modified, both to meet the needs of the sample being tested and to permit use by paraprofessional testers. Therefore, validity of the measures within the PDC environment, and within the test battery in which they are administered, must be determined anew. The concern in this report is with concurrent validity--a measure's relationship to other measures of the same construct and to measures of other constructs; a measure should correlate highly with other measures of the same constructs, and should not correlate at all with measures of independent constructs.

An hypothesized correlation matrix was constructed prior to the fall 1975 data analysis, based on knowledge of the constructs the measures were presumed to represent. The values in the matrix indicate the level of relationship that theoretically should obtain between any two measures if they both genuinely measure the constructs they are supposed to represent. The actual fall 1976 correlations (within language groups) were then evaluated against the hypothesized correlations.

The hypothesized correlation matrix was constructed by first determining the correlation within the three areas of child tests: cognitive-language, psychomotor, and social-emotional measures. Then the desired correlations among the three groups of tests were determined. Generally, higher correlations were expected within an area than between areas. But since each area is actually composed of linked constructs rather than alternative measures of the same construct, very few high correlations were expected.

The fall 1976 correlations between measures (the ones found reliable) were calculated within each language group, and the following procedure was used to determine whether a given measure was valid. First, the obtained intercorrelation matrix was compared with the hypothesized matrix and deviations of each correlation from the hypothesized one were calculated (e.g., if the hypothesized correlation was "medium" and that obtained was "low," a deviation of "-1" was scored; if the hypothesized correlation was "zero" and that obtained was "medium," a deviation of "+2" was scored). For each measure, the absolute values of the deviations were summed across all measures and divided by the number of measures. If this

ratio had a value of 1.0 or less, the measure was considered valid. The criterion implicit in this procedure is that a measure's concurrent validity is adequate if, on the average, the obtained correlations with other measures are within the range adjacent to the expected value.

Although this procedure allows for rather large deviations from the hypothesized relations, it still provides a useful first approximation to validation of the measures.

Step 3: Are Reliability and Validity Constant Across Time and Samples?

This step, illustrated in Figure D-3, Appendix D, is similar to the step described in the last report for ascertaining the fall-to-spring constancy of reliability and validity estimates. Basically, the internal consistency coefficients and validity estimates obtained for each measure for each of the samples were compared and a criterion of constancy was applied to assess the stability of the figures across time (fall 1975, spring 1976, fall 1976) and across samples (Cohort 1, Cohort 2).

The data analysis procedures outlined in this chapter and in previous reports have been used to determine the psychometric characteristics of the child measurement battery for the samples actually tested in 1975-76 (Cohort 1) and 1976-77 (Cohort 2). These analyses also have a broader goal: determining the suitability of this battery for any potential target populations of children that PDC might serve in the future. Since data are now available on two separate samples from this larger population, comparison of the reliability and validity estimates may suggest more generally the appropriateness of the measures as tools for this type of evaluation. Therefore, reliability coefficients and validity estimates from the three available time-points were examined with reference to the question of their constancy. In addition to assessing the constancy of reliability and validity estimates, these analyses may indicate the appropriateness of eventually equating the English and Spanish versions of these measures in order to relate scores on tests given earlier in one language to scores on the same tests given later in the alternate language.

Step 4: Does the Factor Structure Support Expectations?

For both English- and Spanish-dominant samples, a principal components factor analysis was performed on all scale scores found to be reliable for this time-point. After varimax rotation, a table of factor loadings for each language sample (English-dominant, Spanish-dominant) was constituted. Factors were named or described on the basis of the constructs represented by the tests having the highest loadings on each one. Comparability of these fall 1976 results to previous analytic results (fall 1975, spring 1976) was examined and cross-language-sample comparisons were made. The structure of the battery and the stability of that structure were thus determined, and implications were drawn for the future.

Step 5: Are PDC and Comparison Groups Comparable?[1]

PDC and comparison groups were first compared within each site on all the scores produced by the test battery and on variables representing background characteristics. (If data were missing on any background variable for more than 25% of the children in either group, that variable was omitted from the analyses for those groups.) For categorical variables (e.g., ethnicity, sex), equality of PDC vs. comparison group proportions was evaluated by means of χ^2 analyses; the criterion for concluding that a group difference existed was a statistical significance value of less than .01 ($p<.01$). Next, all groups were combined across sites into PDC and comparison aggregates and the same analyses were repeated.

The results of the within-site comparisons generalize to the prospect of conducting site-by-site analyses of program impact in the future; the results of comparisons of the aggregated groups generalize to the prospect of conducting cross-site analyses of impact. Interpretations of the results address the basic question: Are PDC and comparison groups, taken locally or nationally, similar enough to begin with that differences discovered in the future can be attributed confidently to program impact?

---

[1]Georgia, which has no comparison group, was omitted from the within-site analyses described, but not from the analyses of aggregated groups.

Step 6: Are Sample Sizes and Retention Rates Adequate?

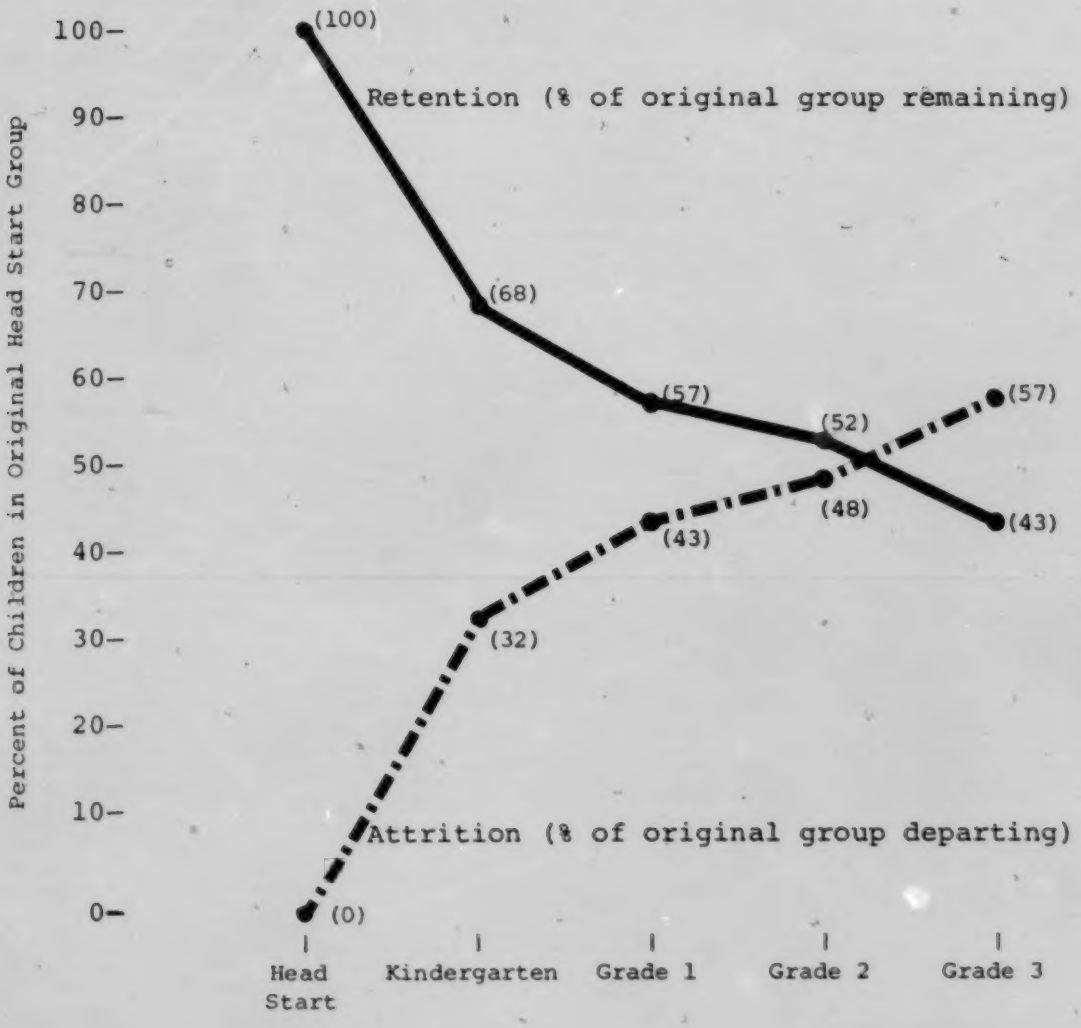
In order to remain in the testing sample from point to point of a longitudinal study, children from the original PDC and comparison Head Start groups must be located in a designated PDC or comparison elementary school, respectively, in each successive year of the study. This is necessary for two reasons: first, the logistics of testing require a central testing location (or no more than a few locations) so that administrative complexities are minimized; second, the Impact Study is basically conceived as a comparative study of two treatments--PDC and non-PDC--and only if children remain in a single treatment can their performance be taken as a reflection of that treatment.

A check was made this fall on the number of children from the 1975-76 testing sample (Cohort 1) who, for the 1976-77 school year, were enrolled in kindergarten at an appropriate school. Since these children presumably come from the same populations as this year's target sample (Cohort 2), the Head Start-to-kindergarten retention rate observed for the earlier sample should provide a sound basis for projecting Cohort 2 retention. However, the retention rate for the first interval (Head Start to K) will not necessarily be the same as the rate for later intervals (K to 1, 1 to 2, 2 to 3). In fact, it generally happens that the greatest loss occurs in the transition from Head Start to kindergarten. Thus the figure representing Head Start-to-kindergarten retention cannot simply be multiplied by itself to project retention in later years. Instead, changes in the rate of retention must be taken into account. For the Cohort 1 retention figures, this was done in the way described in the next paragraph.

Last year and the year before, each site was asked to provide figures from previous years that documented the number of children retained each year in the progression from Head Start through grade 3. Using these figures, a curve was drawn that described the mean observed rate of retention, across all sites, for each grade interval from Head Start through grade 3. This curve, shown in Figure 2, expresses retention rate as the proportion of children from an original Head Start group who can be expected to remain in a designated program (i.e., PDC or comparison) through each successive grade. (The attrition curve also shown in Figure 2 is of

Figure 2

Estimated Retention/Attrition Curve, Head Start through Grade 3, for Head Start Populations at PDC Sites, Based on Data for Groups that Preceded Cohorts 1 and 2



course the mirror image of the retention curve: % attrition = 100 - % retention.) To project future retention of the Cohort 1 children now enrolled in kindergarten, the retention figure obtained for the first interval (Head Start to K) was weighted in such a way as to incorporate the general trend observed for the earlier year groups. Thus no matter what the retention figure for any Cohort 1 group in the Head Start-to-kindergarten interval, the projections for K to 1, 1 to 2, and 2 to 3 describe a curve that parallels the curve for the population. (So, for example, if any group's first-year retention figure was the same as that for the general population, the curve that is extended from that figure will coincide exactly with the general curve.)

These projections provide the basis for forecasting the size of the Cohort 2 PDC and comparison samples in later years of the contemplated longitudinal study. For this step, a table was constructed that began with the actual numbers of children included in this year's PDC and comparison testing samples. Next, for each Cohort 2 group the number was multiplied by the successive retention coefficients obtained for the appropriate Cohort 1 group, adjusted as described. This yielded estimates of the actual number of children likely to remain in PDC and comparison groups, locally and nationally, for each prospective year of the evaluation. The implications of these figures for future statistical analyses of PDC's effects are discussed in the Findings and Conclusions chapter of this report.

III

FINDINGS

Characteristics of the Samples

General Description

In fall 1976, 1,219 children were tested at 12 PDC sites.[1] Table 1 shows the number of children who came from PDC and comparison groups at each site, and describes the makeup of each group in terms of the handicap status, ethnicity, sex, and dominant language of its members.

Not all these children entered into the psychometric analyses presented in this section, but the right-hand column in Table 1 shows the number that did. Children whose dominant language was other than English were excluded from the analytic sample, except in California and Texas, where Spanish-dominant children comprised a sample of their own; also excluded were children with handicaps that were judged likely to impair test performance unduly. Handicap information came principally from local Head Start records. Table 2 gives the frequency of each type of handicap reported among children in the full testing sample. Note that not all of these handicaps were judged to be sufficiently debilitating to warrant a child's exclusion from the analytic sample. It is possible that those who have been excluded at this time will be included in later analyses, but for present purposes it is preferable to restrict the sample to those for whom the measures are likely to possess greatest validity.

---

[1] In the thirteenth site, Arizona, impact was assessed by case study techniques rather than by testing, since testing was found to be unsuitable for the Navajo-speaking children there.

Table 1

Composition of the Samples, Fall 1976

		Number in Full Sample	% Handicapped	ETHNICITY					SEX		DOMINANT LANGUAGE			Number in Final Analytic Sample
				% Black	% Hispanic	% American Indian/ Native Alaskan	% White	% Asian/Pacific Islander	% Male	% Female	% English	% Spanish	% Other	
CALIFORNIA-English	PDC	37	0	3	87	0	11	0	19	81	86	11	3	37
	Comp	25	0	4	63	0	33	0	38	62	96	4	0	24
CALIFORNIA-Spanish	PDC	7	0	0	100	0	0	0	43	57	0	100	0	7
	Comp	15	0	0	100	0	0	0	40	60	0	100	0	15
COLORADO	PDC	55	7	7	75	0	18	0	49	51	100	0	0	51
	Comp	32	6	6	69	0	25	0	56	44	100	0	0	30
CONNECTICUT	PDC	56	11	46	39	2	13	0	59	41	61	36	4	37
	Comp	57	0	84	9	0	7	0	51	49	93	5	2	55
GEORGIA	PDC	46	7	65	0	0	35	0	50	50	100	0	0	43
FLORIDA	PDC	47	4	100	0	0	0	0	64	36	100	0	0	45
	Comp	39	0	87	10	0	3	0	47	53	90	10	0	39
IOWA	PDC	50	2	50	2	0	46	2	50	50	98	0	2	49
	Comp	54	6	24	2	2	70	2	52	48	100	0	0	51
MARYLAND	PDC	44	30	46	9	2	39	5	48	52	94	0	6	32
	Comp	58	22	36	24	0	29	10	52	48	64	17	18	45
MICHIGAN	PDC	66	12	59	5	0	36	0	58	42	100	0	0	58
	Comp	64	9	75	3	0	22	0	45	55	100	0	0	58
TEXAS-English	PDC	26	0	4	31	0	65	0	46	54	100	0	0	26
	Comp	20	5	0	75	0	25	0	50	50	95	5	0	19
TEXAS-Spanish	PDC	38	0	0	100	0	0	0	65	35	0	100	0	38
	Comp	37	8	0	97	0	3	0	41	60	0	100	0	34
UTAH	PDC	68	10	3	15	2	79	2	56	44	99	0	2	61
	Comp	61	10	5	18	8	68	0	49	51	97	2	2	55
WASHINGTON	PDC	58	16	26	0	14	48	12	53	47	97	0	4	49
	Comp	76	13	43	3	5	46	3	42	58	99	0	1	66
WEST VIRGINIA	PDC	46	9	9	0	0	91	0	56	44	100	0	0	42
	Comp	37	22	14	0	0	87	0	38	62	100	0	0	29
TOTALS BY GROUP	PDC	644	9	33	26	2	37	2	53	47	88	11	1	575
	Comp	575	9	36	25	2	36	1	47	53	86	12	2	520
TOTALS, ALL GROUPS COMBINED		1219	9	35	25	2	36	2	50	50	87	11	2	1095

Table 2

Types and Frequencies of Handicaps Reported Among Children in the Full Testing Sample (Total N = 1133)[a]

Type	Frequency	(% of Total)
Physical impairment	12	(1.1)
Hearing impairment	6	(0.5)
Speech impairment	63	(5.6)
Visual impairment	12	(1.1)
Emotional disorder	1	(0.1)
Learning disability	6	(0.5)
Chronic illness	8	(0.7)
Other, non-debilitating[b]	2	(9.7)
All types combined	110	(9.7)

Note. Children with multiple handicaps appear multiply in this table.

[a]This table excludes data from the Florida sample, which were not available at the time of compilation.

[b]A handicap of this type alone was not judged sufficiently severe to warrant a child's exclusion from the testing sample.

Comparability of PDC and Comparison Groups

Once the final analytic samples had been established, analyses were performed to determine just how comparable the PDC and comparison groups really are at each site and in aggregation. Both background characteristics and test performance were examined and the results are shown in Table 3. The background variables examined represent characteristics that have been found in past research to be related to school performance. (If the groups are not initially comparable on these dimensions, it is possible that the effects produced by PDC will be masked by extraneous differences unless these differences are somehow taken into account.)

For each site and for each variable appearing in Table 3, the assumption of PDC-comparison group equality was tested statistically (using the chi-square technique for categorical variables and t tests for metric variables). All available data entered into each analysis, meaning that even if data were missing for a particular child on one or more variables, data obtained for that child on other variables did enter into the respective analyses. A difference was declared to exist between PDC and comparison groups if analysis indicated the chance probability of the observed difference to be less than one in 100 ($p < .01$).

Site-level findings. The asterisks in Table 3 mark statistically significant differences between PDC and comparison groups, the checks mark instances where the groups can be considered comparable or identical.

At the individual site level the groups appear very similar; there are differences on background variables in only one site. On performance measures, of the 13 comparisons made (the Spanish- and English-dominant samples in California and Texas were tested separately in these analyses), ten showed either no group differences or differences only on the POCL or Height and Weight; only two sites had group differences on more than one child measure. For details on the nature of each statistically significant difference found at the site level, see Appendix E.

Aggregate-level findings. For this analysis, data were pooled for English- and Spanish-dominant children separately by treatment group, resulting in four groups: PDC-English, Comparison-English, PDC-Spanish and Comparison-Spanish.[1]

---

[1]Statistics are also shown in Table 3 for an aggregation that excludes Georgia, since Georgia has no contemporaneous comparison group to balance its PDC group in the aggregate. Exclusion of Georgia does not alter the results, however. The combined PDC groups and combined comparison groups appear to be in parallel with or without the Georgia PDC data.

Table 3

Comparability of PDC and Comparison Groups at Each Site and of Groups Aggregated Across Sites, Fall 1976 ($p < .01$)

Key:

* = statistically significant group difference ($p<.01$)
✓ = no significant difference between groups
? = data insufficient for analysis
\- = test not appropriate

		SITE																
		CALIFORNIA-English	CALIFORNIA-Spanish	COLORADO	CONNECTICUT	GEORGIA	FLORIDA	IOWA	MARYLAND	MICHIGAN	TEXAS-English	TEXAS-Spanish	UTAH	WASHINGTON	WEST VIRGINIA	ENGLISH AGGREGATE	SPANISH AGGREGATE	ENGLISH AGGREGATE without Georgia
N for analytic sample	PDC	37	7	51	37	43	45	49	32	58	26	38	61	49	42	530	45	487
	Comp.	24	15	30	55	--	39	51	45	58	19	34	55	66	29	471	49	471
BACKGROUND CHARACTERISTICS																		
Ethnicity		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sex		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Age		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Prior Preschool Experience		✓	✓	✓	✓		✓	✓	✓	✓	✓	*	✓	✓	✓	✓	*	✓
Number of Siblings		?	?	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mother's Education		?	?	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
TEST PERFORMANCE																		
Cognitive-Language Measures																		
BSM-English		✓	?	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	?	✓
BSM-Spanish		-	✓	-	-		-	-	-	-	-	✓	-	-	-	-	✓	-
Block Design (WPPSI)		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Verbal Fluency		✓	✓	✓	✓		✓	*	✓	*	✓	✓	✓	✓	✓	✓	✓	✓
Verbal Memory-1		✓	✓	✓	✓		✓	✓	*	✓	✓	✓	✓	✓	✓	✓	✓	✓
Verbal Memory-3		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Draw-A-Child		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Psychomotor Measure																		
Arm Coordination		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Social-Emotional Measures																		
PIPS-Solutions		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
POCL-Total		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
POCL-1		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
POCL-2		✓	✓	✓	✓		*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Health-Nutrition Measures																		
Height		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	*	✓	✓	✓	✓	✓
Weight		✓	✓	✓	✓		✓	✓	✓	*	✓	✓	✓	✓	✓	✓	✓	✓

No Comparisons Possible

Figures 3 and 4 show the relative standing of the PDC and comparison aggregates on certain background variables and Figures 5 and 6 show their relative standing on 13 performance measures plus three more background variables. As at the site level, the similarities of the aggregated groups are more prominent than their differences. In the English-dominant sample, there are no significant group differences on the background variables and only one difference in test performance. In the Spanish-dominant sample, the groups differed on only one background variable, and there was no difference on any of the performance measures.

Analysis of Attrition Trends

The first column of Table 4 shows, for each site and for all sites collectively, the number of children that were available for fall 1976 testing in PDC and comparison Head Start centers. These children constitute the full sample of Cohort 2, the cohort whose progress will be followed through grade 3 if the PDC evaluation is extended longitudinally. On the average, these groups are about 9% smaller than site staff had estimated they would be (the estimates provided last year by PDC coordinators can be found in Interim Report IV, Volume 1, August 1976). Moreover, the mean retention rate determined this fall for Cohort 1 children (Cohort 2's pilot-year predecessors, now in kindergarten) is lower than was anticipated. To restate these findings on sample size and attrition: Cohort 2 is smaller than it was expected to be, and since it is likely to follow Cohort 1's attrition pattern, its size is likely to diminish faster than was anticipated. These findings are expanded and discussed in the remainder of this section.

The data that had been gathered earlier on past retention rates at each site showed that, on the average, 68% of the four-year-old children enrolled in a given Head Start center went on to enroll in kindergarten at the expected school the following year. Among Cohort 1 children, however, the Head Start-to-kindergarten retention figure is considerably lower--61%. And because this is the figure that provides the basis for projecting the number of Cohort 2 children likely to remain in future years, those projected numbers, too, are smaller than earlier projections.

It is important to note that the number of PDC and comparison children who will actually be available for testing in the future is likely to be even lower than these projections, for a number of reasons. First, the figures shown in Table 4 represent children in the full sample. Although all these children could be tested, consideration of handicap and language factors would

Figure 3

Background Characteristics of Aggregated Analytic Samples of English-Dominant PDC and Comparison Children

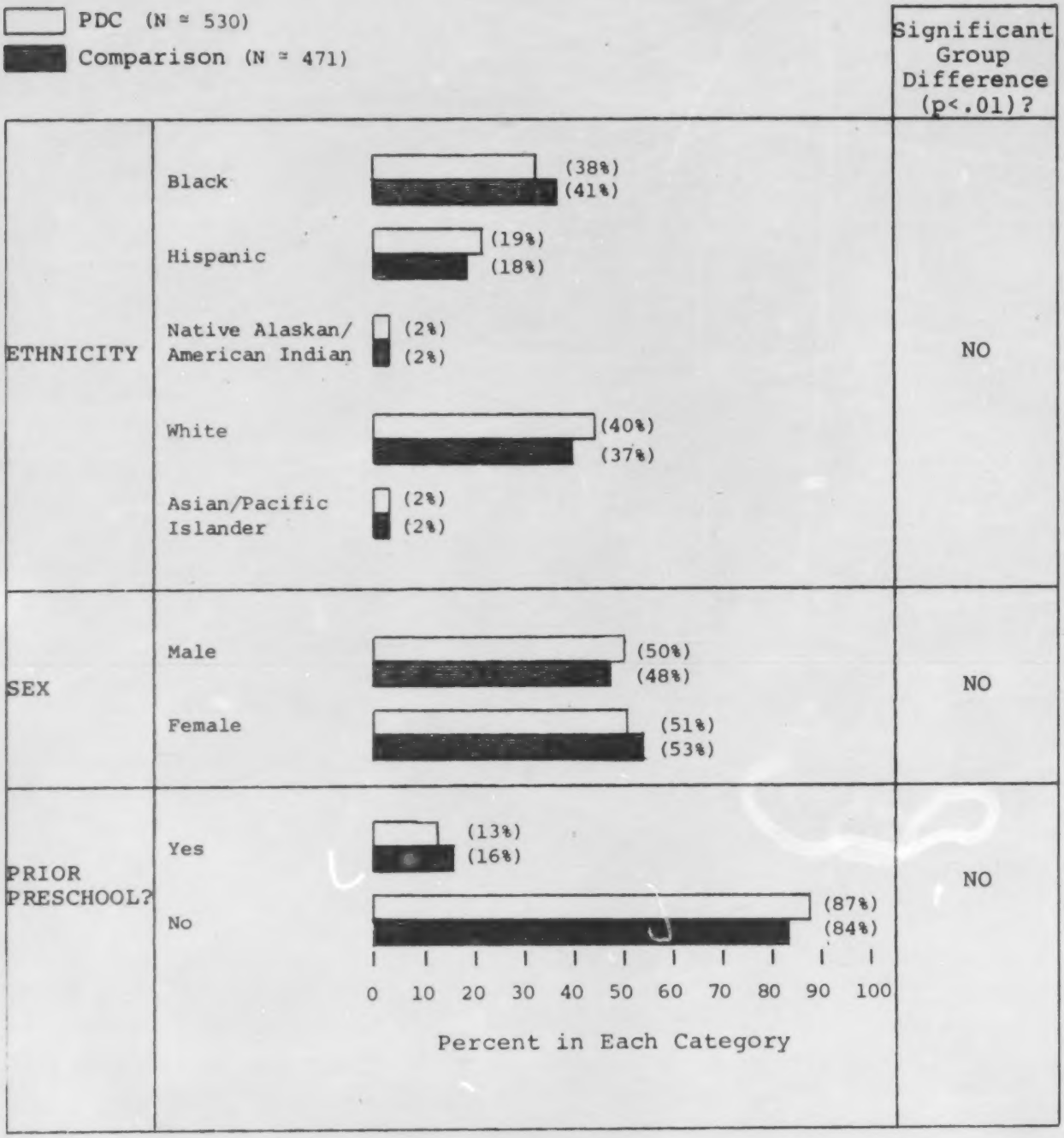


Figure 4

Background Characteristics of Aggregated Analytic Samples of Spanish-Dominant PDC and Comparison Children

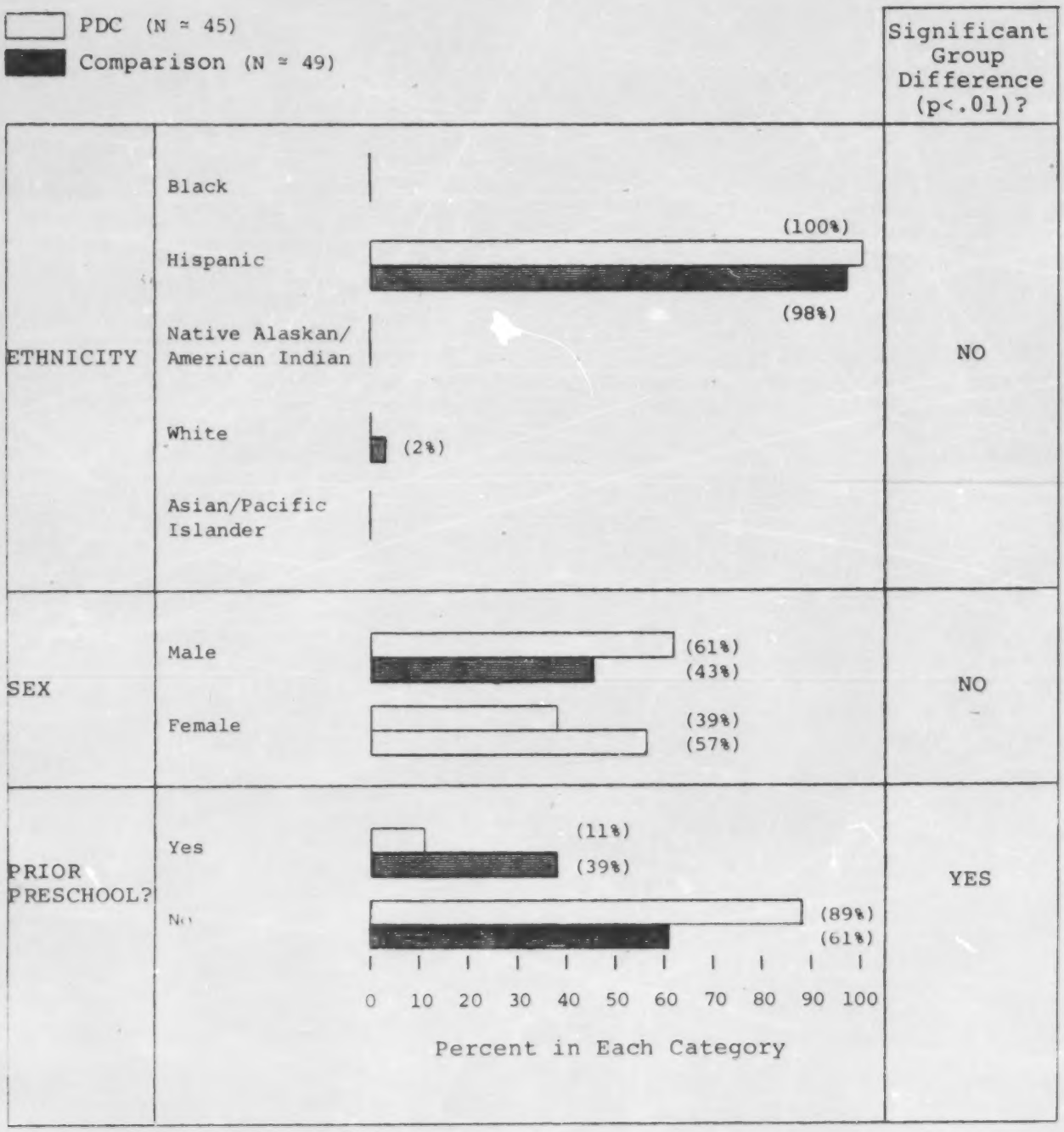


Figure 5

Mean Standard Scores of Aggregated Analytic Samples of English-Dominant PDC and Comparison Children on Selected Demographic and Performance Measures

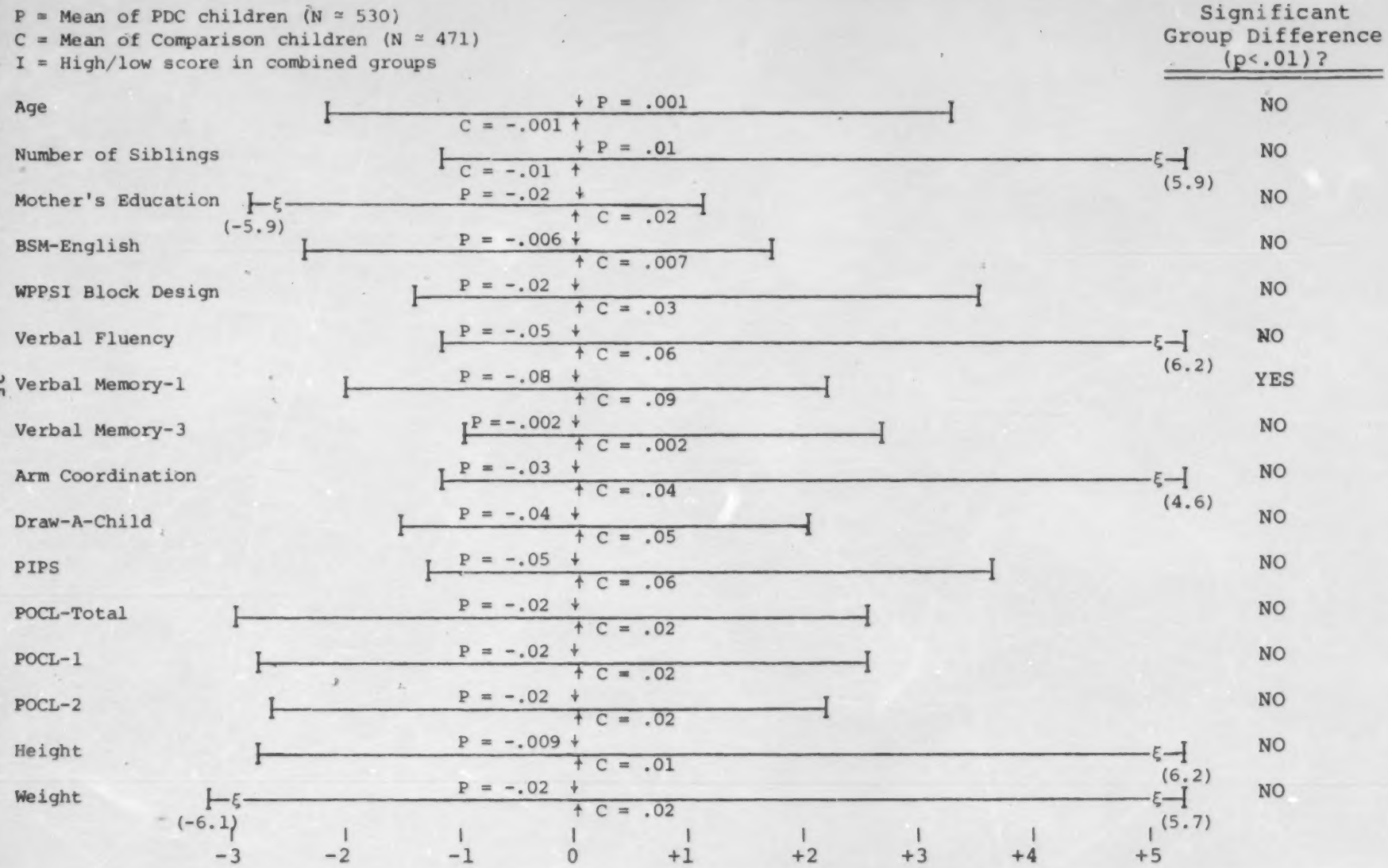


Figure 6

Mean Standard Scores of Aggregated Analytic Samples of Spanish-Dominant PDC and Comparison Children on Selected Demographic and Performance Measures

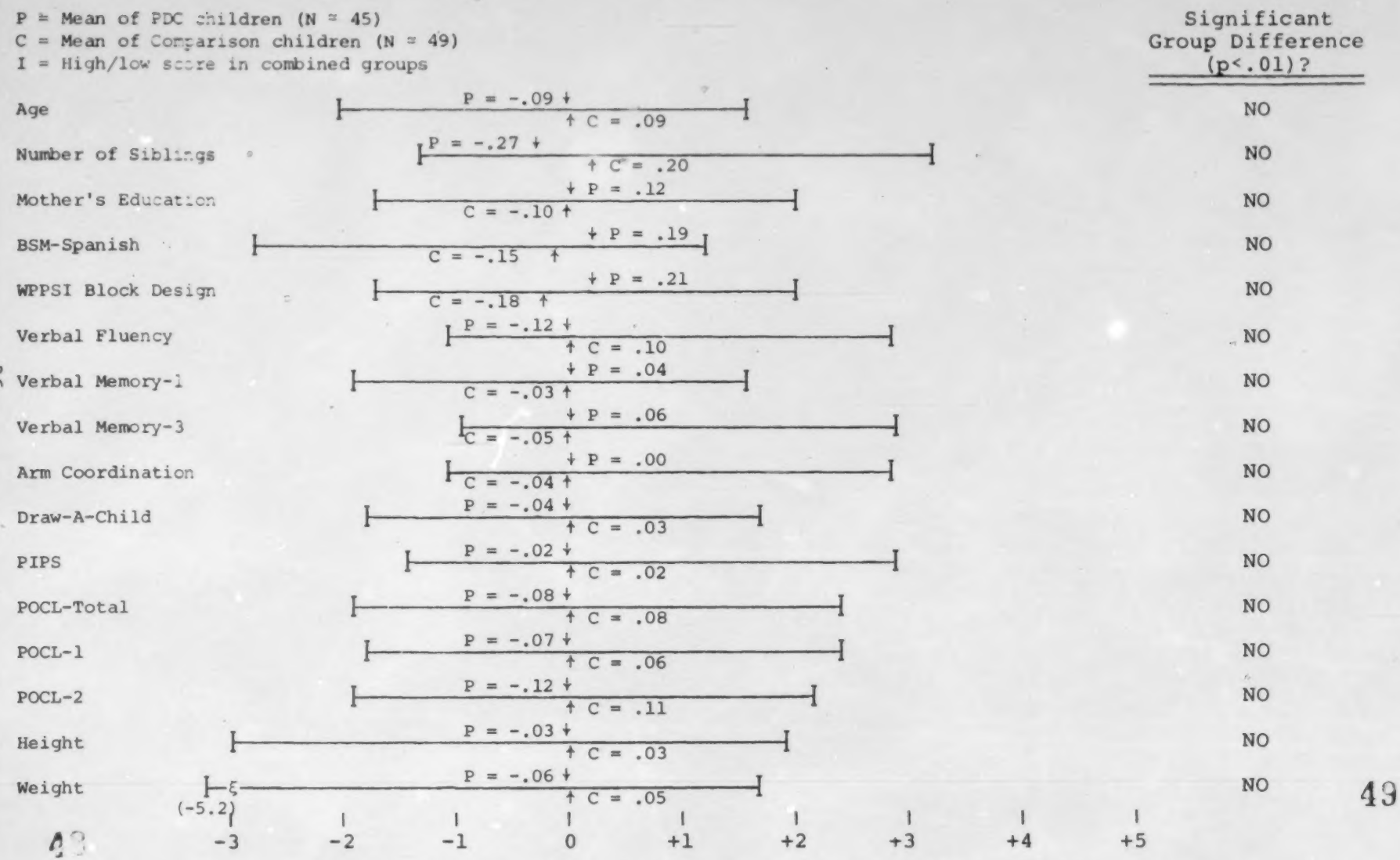


Table 4

Projected Retention of Cohort 2 Children for Each Year of the Prospective Longitudinal Study

		1976-77 Head Start		1977-78 K		1978-79 1		1979-80 2		1980-81 3	
		%	*N*	%	*N*	%	*N*	%	*N*	%	*N*
California	PDC	100	*44*	70	*31*	59	*26*	54	*24*	44	*19*
	Comp	100	*40*	42	*17*	35	*14*	32	*13*	26	*10*
Colorado	PDC	100	*55*	49	*27*	41	*23*	38	*21*	31	*17*
	Comp	100	*32*	42	*13*	35	*11*	32	*10*	26	*8*
Connecticut	PDC	100	*56*	50	*28*	42	*24*	39	*22*	32	*18*
	Comp	100	*57*	42	*24*	35	*20*	32	*18*	26	*15*
Florida	PDC	100	*47*	31	*15*	26	*12*	24	*11*	20	*9*
	Comp	100	*39*	65	*25*	55	*21*	50	*20*	41	*16*
Georgia	PDC	100	*46*	77	*35*	65	*30*	59	*27*	49	*23*
Iowa	PDC	100	*50*	51	*26*	43	*22*	39	*20*	32	*16*
	Comp	100	*54*	38	*21*	32	*17*	29	*16*	24	*13*
Maryland	PDC	100	*44*	73	*32*	61	*29*	56	*25*	46	*20*
	Comp	100	*58*	77	*45*	65	*38*	59	*34*	49	*28*
Michigan	PDC	100	*66*	86	*57*	72	*48*	66	*44*	54	*36*
	Comp	100	*64*	69	*44*	58	*37*	53	*34*	43	*28*
Texas	PDC	100	*64*	84	*54*	71	*45*	65	*42*	53	*34*
	Comp	100	*57*	76	*43*	64	*36*	59	*34*	48	*27*
Utah	PDC	100	*68*	62	*42*	52	*35*	48	*33*	39	*27*
	Comp	100	*61*	36	*22*	30	*18*	28	*17*	23	*14*
Washington	PDC	100	*58*	62	*36*	52	*30*	48	*28*	39	*23*
	Comp	100	*76*	70	*53*	59	*45*	54	*41*	44	*33*
West Virginia	PDC	100	*46*	55	*25*	46	*21*	42	*19*	35	*16*
	Comp	100	*37*	73	*27*	61	*23*	56	*21*	46	*17*
AGGREGATE	PDC	100	*644*	63	*408*	54	*345*	49	*316*	40	*258*
	Comp	100	*575*	58	*334*	49	*280*	45	*258*	36	*209*

NOTE: "%" represents proportion of original group remaining, "*N*" represents size of group remaining. In the 1976-77 column, *N* = original sample size and % = 100, necessarily. The figures in successive columns are projections based on the actual 1976-77 figures.

require the elimination of some children from the analytic sample, which is the source of the data used for statistical analysis. Of the 1,219 children in the full Cohort 2 testing sample, 124--about 10%--were excluded from the analytic sample for reasons of handicap or language. And of the remaining 1,095, 92 comprise the Spanish-dominant analytic sample, leaving 1,001 children (82% of the original 1,219) in the sample that provides the basis for most analyses. Thus, when examining the year-to-year sample size projections shown in Table 4, it should be kept in mind that, in the aggregate, only about 82% of the children who are still present at any point in the future will enter into analyses based on English language test data.

There may, however be some compensations. Some handicaps may be overcome, bringing some of the children who are now excluded because of handicaps back into the analytic sample, and some of the children who are now Spanish-dominant may later become members of the English-dominant sample. But such increments are likely to be offset by other decrements, primarily due to absence from school at testing time and refusal to cooperate--refusal on the part of the child, the child's parents, or school officials. (Of the Cohort 1 children tested in fall 1975, 3% could not be tested the following spring for reasons other than departure from school, and these reasons included absence and refusal.)

Taking these factors into consideration, the number of PDC and comparison children from Cohort 2 who are likely to remain in the analytic sample of English-dominant children through grade 3 can be estimated at about 383 (212 PDC children, 171 comparison children). Estimated similarly, the number of children likely to remain for that period in the Spanish-dominant sample is about 34 (18 PDC, 16 comparison children). What these projections imply for the longitudinal study of PDC's impact is discussed in the Conclusions chapter.

A note on the attrition statistics reported for individual sites: The projections shown in Table 4 can be expected to be more accurate for all sites aggregated than for individual sites. This is because the factors affecting retention within sites may not operate consistently from year to year. For example, a change in school attendance boundaries could mean that a large proportion of a kindergarten-year PDC sample might not return to the same school for first grade. Yet since boundaries are unlikely to be changed every year, retention of the rest of that sample from grade 1 to grade 2 would probably be proportionally greater.

The Florida PDC site was affected by such an occurrence last year. A large fraction of the children from the Cohort 1 PDC Head Start group were assigned unexpectedly to a non-PDC elementary school for their kindergarten year because of over-enrollment in the PDC school. Arrangements have been made at that site to ensure that this will not occur for Cohort 2 children. But it is probable that such events will occur at other sites in coming years (in Iowa's site, for example, a busing plan is under consideration that would alter present school attendance patterns, and this might affect retention of Cohort 2 children at that site). It is assumed that these occurrences will average out across sites, making the aggregate projections in Table 4 more dependable than the projections for individual sites.

Characteristics of the Individually Administered Instruments

Reliability for Cohort 2 in Fall 1976

Table 5 summarizes findings on the reliability of all but two of the instruments included in the fall battery. (The Preschool Interpersonal Problem-Solving Test is not included in this table because its scoring is not amenable to computation of alpha, and the reliability of the PDC Classroom Observation System is treated elsewhere in this chapter.) The coefficients of internal consistency (Cronbach's alpha) were computed separately for the respective analytic samples of English-dominant and Spanish-dominant children, aggregated across sites. As Table 5 shows, all the scales in the English and Spanish test batteries meet the preset reliability criterion--an alpha coefficient of .65 or higher.

Constancy of Reliability Across Time and Cohorts

Table 6 presents comparisons of reliability coefficients (Cronbach's alpha) for the two-cohort, three-time-point data currently available. An inspection of this table suggests an ordering of the measures into three classes: those with fairly constant reliability coefficients, those that show increasing reliability across time-points, and those that show decreasing reliability across time-points.

Measures that show constant reliability. The BSM-English, Block Design (administered in the fall only to both cohorts), Verbal Fluency, Verbal Memory-3, and POCL-Total appear quite stable across time and across cohorts for both the English- and Spanish-dominant samples.

Table 5

Reliability of the Child Measures:[a]
Cronbach's Alpha (Internal Consistency)
for Fall 1976 Head Start Children

Measures	Cronbach's Alpha			
	English-Dominant Children		Spanish-Dominant Children	
COGNITIVE-LANGUAGE	n	r_α	n	r_α
Bilingual Syntax Measure-English	997	.84	39	.93
Bilingual Syntax Measure-Spanish[b]	16	.95	89	.86
Block Design (WPPSI)	999	.77	94	.82
Verbal Fluency (MSCA)	975	.76	92	.81
Verbal Memory-1 (MSCA)	997	.85	94	.89
Verbal Memory-3 (MSCA)	989	.82	93	.84
Draw-A-Child (MSCA)	978	.84	92	.78
PSYCHOMOTOR				
Arm Coordination (MSCA)	976	.65	89	.73
SOCIAL-EMOTIONAL				
POCL-Total (High/Scope)	1001	.95	94	.97
POCL-1 (High/Scope)	1001	.95	94	.96
POCL-2 (High/Scope)	1001	.90	94	.96

[a]Two instruments are not included: the scoring of the Preschool Interpersonal Problem Solving Test does not lend itself to computing alpha, and the reliability of the Classroom Observation System was determined differently.

[b]Texas and California only (Bilingual/Bicultural Demonstration Sites).

Table 6

Comparisons of Reliability (Internal Consistency) for Cohort 1 (Tested Fall 1975 and Spring 1976) and Cohort 2 (Tested Fall 1976)

Measures	Cronbach's Alpha					
	English-Dominant Sample			Spanish-Dominant Sample		
	Fall 1975	Spring 1976	Fall 1976	Fall 1975	Spring 1976	Fall 1976
Bilingual Syntax Measure-English	.82 (691)[a]	.88 (430)	.84 (997)	.93 (17)	.93 (10)	.93 (39)
Bilingual Syntax Measure-Spanish	.88 (13)	--	.95 (16)	.96 (85)	.76 (70)	.86 (89)
Block Design	.75 (724)	.78 (80)	.77 (999)	.80 (87)	--	.82 (94)
Verbal Fluency	.75 (726)	.74 (458)	.76 (975)	.72 (87)	.71 (68)	.81 (92)
Verbal Memory-1[b]	.64 (724)	.73 (435)	.85 (997)	.67 (87)	--	.84 (94)
Verbal Memory-3	.85 (725)	.83 (434)	.82 (989)	.74 (87)	--	.84 (93)
Arm Coordination[c]	.54 (738)	.62 (457)	.65 (976)	.58 (87)	.76 (67)	.73 (89)
Draw-A-Child	.82 (737)	.74 (456)	.84 (978)	.81 (87)	.67 (67)	.78 (92)
POCL-Total	.90 (719)	.93 (462)	.95 (1001)	.87 (87)	.94 (70)	.97 (94)

[a]Numbers in parentheses are sample sizes on which coefficients are based.

[b]Change in content in fall 1976 (see text).

[c]Change in scoring procedure for fall 1976 (see text).

Measures that show increasing reliability. Changes in content and scoring, respectively, on Verbal Memory-1 and Arm Coordination have resulted in higher reliability coefficients for both language samples, taking into account the apparent trend toward increased reliability of the measures in the spring. (That is, internal consistency has tended to be higher in spring than in fall, so in comparing spring 1976 and fall 1976 coefficients, some allowance must be made for this trend.)

Measures that show decreasing reliability. Only one scale, Draw-A-Child, falls into this class. As noted in the August 1976 Impact Study Report, most children obtain near-maximum scores on this measure as they approach ages 5 and 6; and as variance in item and total scores diminishes, so must the reliability coefficient. An alternative to the standard McCarthy scoring procedure is available that is more sensitive to the finer differences that appear among the drawings produced by older children, and this procedure will be explored as a means of restoring reliability of the measure for children at levels higher than Head Start.

Validity

All of the instruments whose reliability was examined in the preceding sections show acceptable evidence of validity for use with Head Start children, as discussed below.

The validation procedures (described more fully in the Methods section of this report) involved determining the expected relationship of each measure with each other one, then comparing these expectations with the relationships that actually appeared in the data. Under this convergent-discriminant method of assessing validity, the assumption is made that if an instrument is actually measuring the construct it was intended to measure, the results will correlate highly with other measures of the same general construct, will correlate moderately with measures of similar constructs, and will not correlate at all with measures of independent constructs. Table 7 displays the matrix of expected relationships. Tables 8 and 9 contain the actual correlations for English- and Spanish-dominant samples of Head Start children, combined across groups and sites, for the fall 1976 testing period. (Note that the actual correlations are presented for some measures that do not appear in the hypothesized matrix--POCL subscales, height, and weight--because no expected relationship was stated.)

Table 7

Hypothesized Correlation Matrix
for PDC Child Measures

	CHILD MEASURES	BSM-English/ Spanish	Block Design	Verbal Fluency	Verbal Memory-1	Verbal Memory-3	Draw-A-Child	Arm Coordination	PIPS	POCL-Total
COGNITIVE-LANGUAGE	BSM-English/Spanish	1.00								
	Block Design (WPPSI)	Med	1.00							
	Verbal Fluency	Med	Low	1.00						
	Verbal Memory-1	Med	Low	Med	1.00					
	Verbal Memory-3	Med	Low	Med	Med	1.00				
	Draw-A-Child	Low	Med	Low	Low	Low	1.00			
PSYCHO-MOTOR	Arm Coordination	0	0	0	0	0	Low	1.00		
SOCIAL-EMOTIONAL	PIPS	Low	Low	Low	Low	Low	Low	0	1.00	
	POCL-Total	Med	Med	Med	Med	Med	Med	Low	Med	1.00

Scale

O: -.1 to .1
Low: .1 to .3
Med: .3 to .5
High: .5+

Table 8
Intercorrelations of Child Measures for English-Dominant Sample[a]
PDC Fall Data, 1976

CHILD MEASURES		BSM-English	Block Design	Verbal Fluency	Verbal Memory-1	Verbal Memory-3	Draw-A-Child	Arm Coordination	PIPS	POCL-Total	POCL-1	POCL-2	Weight	Height
COGNITIVE-LANGUAGE	BSM-English	1.00												
	Block Design (WPPSI)	.28 (912)	1.00											
	Verbal Fluency	.28 (892)	.29 (894)	1.00										
	Verbal Memory-1	.28 (910)	.16 (913)	.29 (892)	1.00									
	Verbal Memory-3	.31 (902)	.25 (905)	.41 (885)	.44 (905)	1.00								
	Draw-A-Child	.32 (893)	.44 (895)	.34 (895)	.20 (893)	.27 (886)	1.00							
PSYCHO-MOTOR	Arm Coordination	.02 (890)	.14 (892)	.13 (892)	.08 (890)	.08 (884)	.04 (883)	1.00						
SOCIAL-EMOTIONAL	PIPS	.31 (893)	.20 (895)	.37 (895)	.28 (893)	.34 (886)	.23 (896)	.06 (893)	1.00					
	POCL-Total	.23 (913)	.19 (915)	.37 (895)	.27 (913)	.31 (905)	.23 (896)	.13 (893)	.36 (896)	1.00				
	POCL-1	.21 (913)	.22 (915)	.36 (895)	.27 (913)	.31 (905)	.26 (896)	.14 (893)	.36 (896)	.97 (916)	1.00			
	POCL-2	.21 (913)	.09 (915)	.30 (895)	.20 (913)	.24 (905)	.13 (896)	.08 (893)	.29 (896)	.83 (916)	.68 (916)	1.00		
HEALTH	Weight	.05 (901)	.12 (903)	.15 (890)	.11 (901)	.10 (893)	.04 (891)	.15 (888)	.11 (891)	.10 (904)	.09 (904)	.08 (904)	1.00	
	Height	.03 (899)	.14 (901)	.16 (888)	.10 (899)	.12 (891)	.08 (889)	.16 (886)	.13 (889)	.06 (902)	.05 (902)	.04 (902)	.70 (902)	1.00

[a]Sample size for each correlation is shown in parentheses.

Table 9

Intercorrelations of Child Measures for Spanish-Dominant Sample[a]
PDC Fall Data, 1976

	CHILD MEASURES	BSM-Spanish	Block Design	Verbal Fluency	Verbal Memory-1	Verbal Memory-3	Draw-A-Child	Arm Coordination	PIPS	POCL-Total	POCL-1	POCL-2	Weight	Height
COGNITIVE LANGUAGE	BSM-Spanish	1.00												
	Block Design (WPPSI)	.28 (89)	1.00											
	Verbal Fluency	.33 (88)	.38 (92)	1.00										
	Verbal Memory-1	.42 (89)	.37 (94)	.43 (92)	1.00									
	Verbal Memory-3	.32 (88)	.27 (93)	.43 (91)	.42 (93)	1.00								
	Draw-A-Child	.24 (88)	.42 (92)	.29 (92)	.39 (92)	.17 (91)	1.00							
PSYCHO-MOTOR	Arm Coordination	-.09 (86)	.18 (89)	.09 (89)	.12 (89)	-.12 (88)	.04 (89)	1.00						
SOCIAL-EMOTIONAL	PIPS	.42 (88)	.16 (92)	.35 (92)	.25 (92)	.33 (91)	.11 (92)	-.05 (89)	1.00					
	PoCL-Total	.29 (89)	.12 (94)	.25 (92)	.32 (94)	.19 (93)	.14 (92)	.10 (89)	.37 (92)	1.00				
	POCL-1	.31 (89)	.16 (94)	.28 (92)	.36 (94)	.21 (93)	.17 (92)	.12 (89)	.38 (92)	.99 (94)	1.00			
	POCL-2	.24 (89)	.02 (94)	.16 (92)	.20 (94)	.11 (93)	.07 (92)	.04 (89)	.30 (92)	.93 (94)	.88 (94)	1.00		
HEALTH	Weight	.00 (85)	.20 (90)	.19 (88)	.17 (90)	.17 (89)	.01 (88)	.10 (86)	.21 (88)	.03 (90)	.08 (90)	-.11 (90)	1.00	
	Height	.22 (85)	.26 (90)	.23 (88)	.17 (90)	.12 (89)	.23 (88)	.14 (86)	.20 (88)	.03 (90)	.09 (90)	-.12 (90)	.54 (90)	1.00

[a]Sample size for each correlation is shown in parentheses.

The more consistent the expected and the actual relationships are for any measure, the stronger the basis for assuming measurement validity. Under this procedure, degree of validity is expressed in terms of the mean absolute deviation of expected from actual relationships. For example, if it was expected that the correlation of Measure A with Measure B would fall in the range defined as "medium" but it actually fell in the low range, this represented a deviation of 1 unit from the expectation. Deviations were summed and averaged to produce a validity index for each measure. The deviations for the current testing period are shown in Tables 10 and 11.

Figure 7 presents profiles of the indexes so obtained for each measure at each of the three testing periods (fall 1975, spring 1976, and fall 1976) for the English- and Spanish-dominant samples of Head Start children. A mean deviation of 1 or less was accepted as evidence of validity (deviations could theoretically range from 0 to 6). According to this criterion, all the instruments examined are acceptably valid for Head Start children, as evidenced by the stability of their validity indexes across two cohorts and three time-points.

Characteristics Determined from Past Reporting Periods: Sensitivity to Change, Suitability for Older Children, and Relationship to Social Competence

The preceding sections of this report have dealt with the reliability and validity of the child measures based upon results of fall 1976 testing and upon earlier PDC testing. Although the fall 1976 data cannot be used at this time to examine two other critical test characteristics (sensitivity to change and suitability for use with older children), it is appropriate at this point to review what we have learned about these characteristics from the 1975-76 testing periods.

Since the Impact Study will depend upon the PDC battery of measures to detect change that can be attributed to program differences, several analyses based on fall 1975 and spring 1976 data were devoted to determining the extent to which the measures included in the battery are sensitive to change. The results of those analyses were considered to be only estimates of the sensitivity that the measures are likely to show in the future. (In fact, the results are probably under-estimates because of a short fall-to-spring interval and some test revisions that took place between testing periods.)

Table 10

Deviations of Child Measure Correlations from Hypothesized Correlations, English-Dominant Sample
PDC Fall Data, 1976

	CHILD MEASURES	BSM-English	Block Design	Verbal Fluency	Verbal Memory-1	Verbal Memory-3	Draw-A-Child	Arm Coordination	PIPS	POCL-Total
COGNITIVE-LANGUAGE	BSM-English									
	Block Design (WPPSI)	-1								
	Verbal Fluency	-1	0							
	Verbal Memory-1	-1	0	-1						
	Verbal Memory-3	0	0	0	0					
	Draw-A-Child	+1	0	+1	0	0				
PSYCHO-MOTOR	Arm Coordination	0	+1	+1	0	0	-1			
SOCIAL-EMOTIONAL	PIPS	+1	0	+1	0	+1	0	0		
	POCL-Total	-1	-1	0	-1	0	0	-1	0	

KEY

0: Correlation was within hypothesized range
+1 to +3: Correlation was specified number of levels higher than hypothesized
-1 to -3: Correlation was specified number of levels lower than hypothesized

Table 11

Deviations of Child Measure Correlations from Hypothesized Correlations,
Spanish-Dominant Sample
PDC Fall Data, 1976

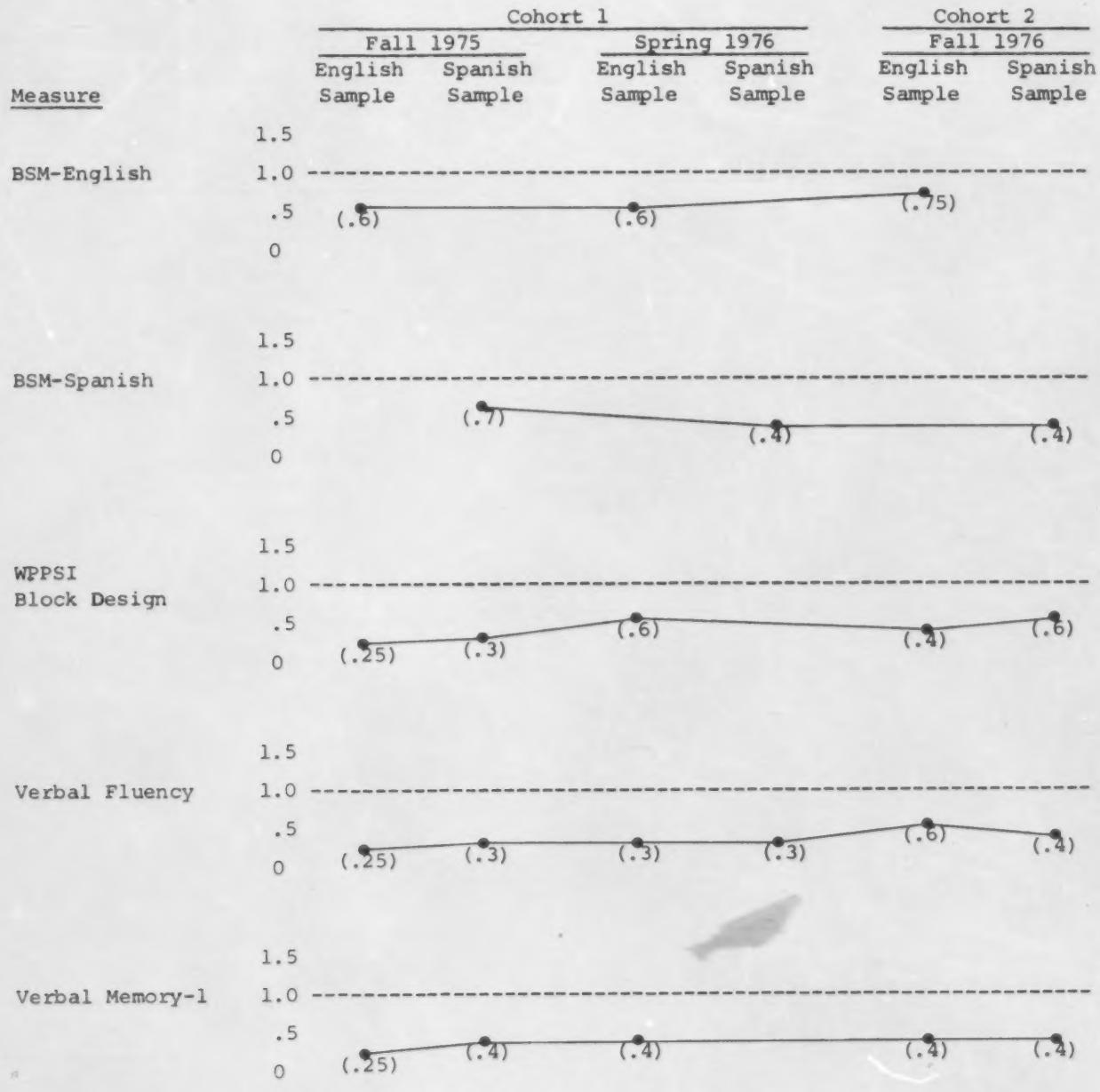
	CHILD MEASURES	BSM-Spanish	Block Design	Verbal Fluency	Verbal Memory-1	Verbal Memory-3	Draw-A-Child	Arm Coordination	PIPS	POCL-Total
COGNITIVE-LANGUAGE	BSM-Spanish									
	Block Design (WPPSI)	-1								
	Verbal Fluency	0	+1							
	Verbal Memory-1	0	+1	0						
	Verbal Memory-3	0	0	0	0					
	Draw-A-Child	0	0	0	+1	0				
PSYCHO-MOTOR	Arm Coordination	0	+1	0	+1	-1	-1			
SOCIAL-EMOTIONAL	PIPS	+1	0	+1	0	+1	0	0		
	POCL-Total	-1	-1	-1	0	-1	-1	-1	0	

KEY

0: Correlation was within hypothesized range
+1 to +3: Correlation was specified number of levels higher than hypothesized
-1 to -3: Correlation was specified number of levels lower than hypothesized

Figure 7

Validity Profiles for Child Measures, for two Head Start Cohorts at a Total of Three Time Points

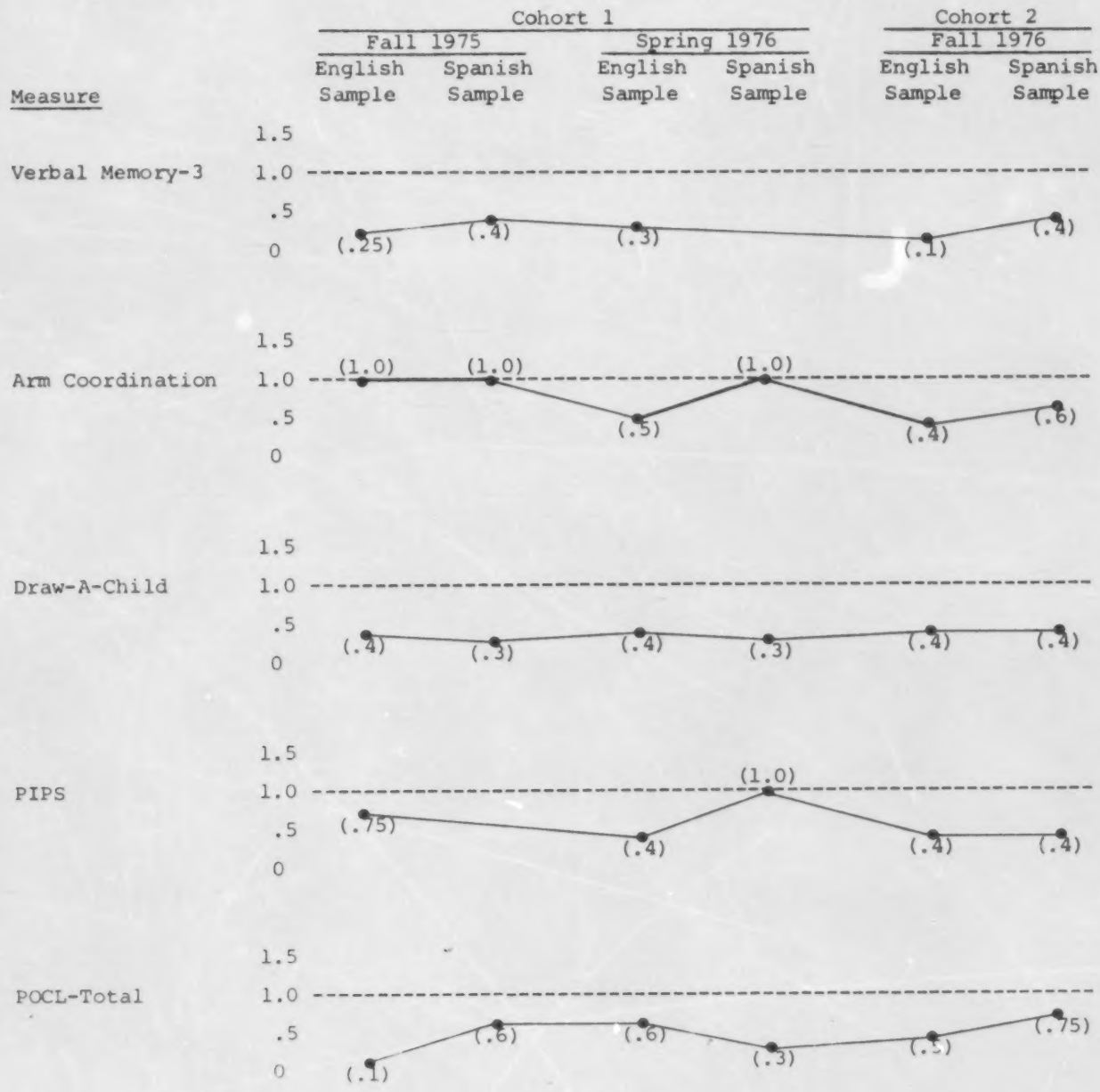


Note. The points plotted represent the mean absolute deviation of the measure from expected relationships with other measures. The broken line represents the level (1.0) above which deviations are considered excessive. (The range of theoretically possible deviations extends from 0 to 6.)

continued:

Figure 7

Validity Profiles for Child Measures, for two Head Start Cohorts at a Total of Three Time Points (continued)



The sensitivity-to-change analyses were of three types. First, the correlation of each measure, in the fall and in the spring, with child age at the time of testing was calculated to determine the age-relatedness of the measures (a substantial relationship was expected for most measures). Next, a paired (or correlated) t test was calculated for the difference between the fall mean score and the spring mean score on each measure to ascertain if the scores increased significantly from fall to spring. Finally, a regression procedure was used to determine whether the observed spring mean on a measure was equal to or greater than the expected, or predicted, spring mean. This procedure would ascertain whether children gained at least as much as they were expected to gain over the given time interval. The results of the analyses were more critical for some measures than for others. For example, scores on the Child Rating Scale and the POCL were expected to be less related to age than were scores on the other measures, since the ratings were being made relative to other children in the same age group. And while these ratings were expected to change from fall to spring, they were not expected to change in a consisted up-or-down direction.

Correlations with age. The correlations of each of the measures with child age at the time of testing tended to be low, positive, and significant, with coefficients generally around .15 to .30. They had not, in fact, been expected to be much higher, since the measures would not be useful for a program evaluation if they were related only to age rather than to differential experience. The only non-significant correlations (other than for WPPSI Block Design, see below) were within the Spanish-dominant sample, and although these correlations were not substantially lower than they were in the English-dominant sample, they were not statistically significant due to the smaller size of the Spanish-dominant sample.

WPPSI Block Design showed a small negative correlation with age. This correlation was not expected to be negative, but neither was it expected to be significantly positive, because it is presumed to be a measure of general ability, a trait that is likely to remain invariant over short intervals. No other sensitivity-to-change tests could be performed for Block Design because it was given at only one time-point.

Fall-to-spring change: t tests. The Bilingual Syntax Measure-English showed a non-significant increase for Spanish-dominant children. The sample sizes, however, were extremely small (5 and 4); so the measure may or may not be sensitive to change (at least for this short a time period) when the language of the test is not the child's primary language. All other child measures showed a significant fall-to-spring increase.

Fall-to-spring change: regression analysis. For four of the child measures (BSM-E, Draw-A-Child, Verbal Fluency, and Verbal Memory-3), R^2 values (representing predictive power) were comparable for fall and spring regression equations which related child status variables (age, sex, ethnicity, preschool experience, and siblings) to child score on the measure. For each of these four measures, therefore, an expected spring score was calculated based on fall score, status on background variables, and spring age. For Draw-A-Child, Verbal Fluency, and Verbal Memory, the actual spring mean was greater than the expected spring mean; i.e., the children gained more on the measures than was expected as a function of their increase in age. For all four of the measures, more than half of the children obtained an actual spring score that was equal to or greater than their expected spring score. These results imply that the tests are sensitive to change that is due to educational experience in addition to experience that is simply a function of increased age. The fact that this analysis was not performed for the other measures does not mean that those tests are not sensitive to such change; they were excluded because the assumptions underlying the calculation of the predicted spring score did not appear tenable for those particular tests.

Summary of sensitivity to change. Based on the results of the three analyses discussed above, it was concluded that all of the child measures being used in the 1976-77 PDC evaluation are in fact sensitive to change, within the developmental range represented in the 1975-76 Head Start sample.

Suitability of the instruments for use in the higher grades. Part of the task of assessing the usefulness of the child measures for the PDC evaluation has been to determine how suitable they for use with children in kindergarten through grade 3 as well as for use with Head Start children. During the 1975-76 testing periods, approximately 25 children per grade (kindergarten through grade 3) were tested in the Georgia site as part of the cross-sectional design there. In addition, 30 third graders were tested in Maryland. This information was used to determine how suitable the instruments can be expected to be for future use at those grade levels, as will be necessary if the evaluation is extended.

Conclusions about the suitability of the child measures for use in kindergarten through grade 3 were based on four factors, which were considered at each grade level: response distributions on the items of each measure, mean scores on each measure, reliability (internal consistency) of the measures, and validity of the measures. The criteria of acceptability for these four factors were as follows:

- Response distributions: An item was considered not to be useful at a given grade level if more than 75% of the children at that level received the maximum score for the item.

- Mean scores: Mean scores should increase systematically across age levels, except for the rating scales. In addition, if the mean score at any level was greater than 80% of the total possible score for the test, the measure was considered unsuitable for children of that age.

- Reliability and validity: The criteria for these two factors were essentially the same as those applied to findings for the Head Start sample discussed previously.

Based on these four factors, most of the measures appear to be useful through grade 3, either in their present forms or with modification. Each child measure is dicussed below.

- Arm Coordination: All six items were useful, based on response distributions, from Head Start through grade 2, and only two items were not useful at grade 3. The mean scores increased acceptably across age levels and did not reach a ceiling on the measure. The only indication of difficulty with this measure is in the validity profile, where the deviation from expected correlation levels is rather large at grade 2. However, the deviations appear to peak at grade 2 for all of the measures, so this sample may not accurately represent the general population. Arm Coordination is therefore judged to be suitable for all grade levels.

- BSM-English and BSM-Spanish: These measures will probably be suitable for the evaluation through grade 3. Intermediate grades were not tested on these measures, so not enough information is available to make a definite decision regarding them. The mean scores and validity profiles are acceptable. The number of non-useful items (12 or 18 at grade 3 for the English version) and the third-grade internal consistency (alpha = .58) appear to be problematic.

- Draw-A-Child: Scores on this measure approach the maximum by grade 1, making it unusable in its present form beyond kindergarten. The problem with this test, however, appears to be in the scoring rather than in the nature of the task. We have reduced the complexity of the scoring from that recommended by the test author.

We would now recommend returning to the more complex scoring or to a Goodenough-Harris type of scoring in order to make use of the valuable information that the drawings of older children can provide. With a change in scoring procedures, Draw-A-Child is expected to be suitable for use across the age range that will be spanned by a longitudinal study of PDC.

- PIPS: Based on all four considerations, the PIPS is judged to be suitable for use at all grade levels of the evaluation.

- POCL and Child Rating Scale: Both rating scales are judged to be suitable for use at all grade levels. Mean scores, alphas, and validity appear to be acceptable. The response distributions are acceptable, but tend to be unusually centrally distributed for the POCL for grades 1 through 3 (which may be attributable to the particular testers who assigned the ratings--this will be investigated in the future).

- Verbal Fluency: Based on all four considerations, Verbal Fluency is judged to be suitable for use at all grade levels.

- Verbal Memory-1: This test produces scores very near ceiling at Head Start, and last spring was found to be unusable beyond kindergarten because most of the children received the maximum score. Accordingly, for fall 1976 testing two items were added to the original four; the added items were constructed on the models of the original word strings, but being longer, present a more difficult task. Logically, this should result in a test ceiling that will be reached by fewer older children, but this proposition has not been tested. Among this year's Head Start children, though, the internal consistency of this scale was higher than it was for last year's Head Start children.

- Verbal Memory-3: Based on all four considerations, Verbal Memory-3 is judged to be suitable for use at all grade levels to be spanned by the evaluation.

Relationship to social competence. Since the PDC battery was constituted with the intent of providing for measurement of the traits that comprise social competence, an analysis was performed for the last Impact Study report that examined the relationship of spring 1976 test scores (for Cohort 1 and higher-grade

samples) to ad hoc criteria of social competence. The criteria were established by factor analyzing ratings from the PDC Child Rating Scale and the Pupil Observation Checklist (POCL)--instruments completed by each child's teacher and tester respectively--then, on the basis of the results, creating factor scores for each child that represented his or her status on each of the "social competence" factors. The Child rating Scale and POCL were chosen as sources of the proxy criteria because the assessments provided by the teacher and testers are based upon observations of each child's behavior in a variety of formal and informal situations, and thus logically come close to representing measures of the child's "everyday effectiveness," i.e., social competence.

The object of the analysis (a linear regression procedure) was to determine the magnitude of the relationship existing between the tests included in the PDC battery and the "social competence" criteria. The more relevant the tests are to social competence, the stronger the relationship expected. The tests that entered into the analysis included BSM-English, Arm Coordination, Draw-A-Child, Verbal Fluency, Verbal Memory-1, Verbal Memory-3, and the PIPS. All of these except Arm Coordination were found to be substantially associated with the collective "social competence" criteria: R^2 values ranged from .17 for BSM-English to .37 for the PIPS; R^2 for Arm Coordination was .05. For children in the higher-grade sample (consisting of pooled K-3 samples) a significant relationship was found for all the tests except Draw-A-Child (whose low reliability for older children has already been noted) and Verbal Memory-1 (which, before this year's revision, also showed low reliability among older children). Arm Coordination was significantly related to the "social competence" criteria for this sample (R^2 = .12), perhaps because of the greater variance occurring across the broader age range. Verbal Memory-3 and the PIPS showed the strongest relationship to the predictors for this sample; both had R^2 coefficients of .25.

This technique, although quite exploratory, produced results that constitute at least preliminary confirmation of expectations: all the aforementioned tests, originally selected for their theoretical relevance to social competence, seem to provide measures that are empirically relevant to social competence, or to our best approximation of that construct.

Ease of Administration

It is of little use to select tests which may have all of the necessary psychometric qualities if they cannot be administered satisfactorily by the testers hired in each PDC community. Therefore, one of the factors taken into consideration when tests were being reviewed for the PDC Impact Study was their general suitability for administration by a paraprofessional. Even though all of the tests selected for PDC met this requirement, there were some differences in the ease/difficulty of their administration. The details presented below are based on observations of tester performance during the tester training session and on the continual feedback by testers throughout the data collection period. (See Appendix B for the monitoring forms used as checks on compliance with testing specifications.)

Bilingual Syntax Measure (BSM). This test is relatively easy to administer--the directions are straightforward and the cartoonlike pictures usually capture the children's attention and elicit responses to the items. The only problematic area of the BSM administration is recording the child's response accurately. Testers have to listen carefully to the child's answer and record it verbatim. This has to be a conscious effort since adults tend to write a correct verb tense or word ending automatically when the child has used incorrect tense or wording.

Verbal Memory. This test presents few administration problems. In Part 1 the tester slowly reads a string of words and records those the child repeats, while in Part 3 (Part 2 is not included in the PDC battery) the tester reads a story and records the child's account of it. The only administration error noted, a minor one, is a tendency on the part of some testers to paraphrase the standard encouragements rather than repeat them exactly.

WPPSI Block Design. More time was spent practicing this test than any of the others because of the 10 different designs the tester must learn to construct with the blocks, and because of the lengthy instructions. The tester has to learn to make the designs while simultaneously reading aloud the instructions for constructing the design. However, once familiar with construction of the designs, testers can usually coordinate both activities. Thereafter, only minor wording errors tend to occur during the actual test administration.

Draw-A-Child. No problems were encountered in the administration of this test. Most of the children enjoyed it and it was a good lead-in test for the second session.

Verbal Fluency. This test is easy to administer. The tester instructs the child to name all of the toys (or things to eat or names of people, etc.) he can within a specific time period. Only a minor problem was discovered--a tendency among testers to paraphrase the encouragements.

Preschool Interpersonal Problem-Solving Test (PIPS). Most of the administration problems encountered with the PIPS were related to scoring. The object of the test, which presents the child with hypothetical interpersonal problems, is to have the child suggest as many different solutions as he can think of (e.g., If A has the truck and B wants to play with it, what can B do to get a chance to play with it? If C has the sailboat and D wants to play with it, what can D do to get a chance to play with it?). Once the child responds, the tester has to decide whether or not the response is relevant and whether or not the child has given a similar answer already. If the response is new and relevant, the tester must decide in which of 16 scoring categories to place the answer (e.g., ask, share, trade, bribe). Many of the responses that children give are clearly understandable such as, "ask him for it," "share it," "tell his mother," but other answers require more judgment on the part of the tester and, as a result, some of these are scored incorrectly. Therefore, testers need to have a clear understanding of each category and how it differs from the others.

Arm Coordination. Although appearing simple and straightforward, this test has the greatest potential for presenting difficulties in administration since the tester has to attend to so many details at the same time. For example, in Part 2 (beanbag catch game) the tester must coordinate reading of the instructions with tossing the beanbag while also watching to see if the child stepped over the line and if he used the correct hand to catch the beanbag. The tester then must record the child's response while making sure the child isn't tossing the beanbag to her. Since the tester has to attend to many details during the test it is easy to commit wording errors or to overlook some of the necessary activities.

Pupil Observation Checklist (POCL). Some testers had problems in differentiating among children when completing this rating scale. That is, some tended to rate every child "average" on all of the dimensions. For example, even though the instructions indicated that a child should be rated "average" on the "Cooperative-Resistive" dimension if the child was as cooperative or as resistive as other children during testing, it seems unlikely that all children would behave identically during the test sessions. More specific rating scale instructions may be needed for those testers who tended to rate all children the same.

Summary. Generally, then, the tests are not difficult to administer. Tester performance improves with practice and the administration difficulties mentioned in this section are more apparent with new testers than with experienced ones.

Factor Structure of the Battery

Factor analyses[1] were performed to investigate inter-relationships among the child measures in the fall 1976 battery. Separate analyses were performed for the English-dominant and Spanish-dominant samples, with the results shown in Tables 12 and 13. While parallel versions of the same measures were included in the two analyses, the substantial difference in sample sizes (English = 880, Spanish = 85) constrains the expected comparability of the results. In view of that limitation, the resulting factor structures of the total battery for the separate samples are remarkably similar.

Results for the English-dominant sample. Analysis yielded three distinct factors which, after rotation, accounted for 26.7%, 19.3%, and 11.6% of the variance, respectively. The scales that loaded highest on the first factor, which could be labeled "verbal-responsive," were Verbal Fluency, Verbal Memory-1, Verbal Memory-3, PIPS, and POCL.[2] The BSM-English also loaded substantially on this factor, but loaded higher yet on the second factor. The scales that loaded highest on factor 2, which could be called "cognitive flexibility," were BSM-English, WPPSI Block Design, and Draw-A-Child. The third factor appears to represent a "psychomotor" dimension. Only Arm Coordination had a substantial loading on this factor. All three factors combined accounted for 57.6% of the variance among measures for the English-dominant sample.

Results for the Spanish-dominant sample. Factor analysis yielded three distinct factors which, after rotation, accounted for 13%, 26.1% and 22.3% of the variance, respectively. On the first factor, apparently representing the dimension "psychomotor," only Arm Coordination had a substantial loading. On the second factor, which might be designated "verbal-cognitive," the scales that had the highest loadings were WPPSI Block Design, Verbal Fluency, Verbal Memory-1, Verbal Memory-3, and Draw-A-Child. On the third factor, which could be termed "verbal-resonsive," the tests loading highest were BSM-Spanish, PIPS and POCL. Combined, these three factors accounted for 61.4% of the variance among measures for the Spanish-dominant sample.

---

[1] Principal components solution, varimax rotation.

[2] A factor analysis of the POCL itself is presented in Appendix G.

Table 12

Factor Analysis[a] of Scores on Child Measures,
English-Dominant Head Start Children,
Fall 1976 Data

N=880

	Factor Loading of Child Measures (highest loading italicized)		
CHILD MEASURE	Factor 1	Factor 2	Factor 3
BSM-English	.41	*.48*	.20
WPPSI Block Design	.09	*.82*	-.17
Verbal Fluency	*.59*	.36	-.15
Verbal Memory-1	*.70*	.04	.01
Verbal Memory-3	*.71*	.19	.01
Arm Coordination	.08	.05	*-.95*
Draw-A-Child	.17	*.80*	.03
PIPS	*.67*	.14	.03
POCL	*.64*	.11	-.19

[a]Principal components solution, varimax rotation.

Table 13

Factor Analysis[a] of Scores on Child Measures,
Spanish-Dominant Head Start Children,
Fall 1976 Data

N=85

	Factor Loading of Child Measures (highest loading italicized)		
CHILD MEASURE	Factor 1	Factor 2	Factor 3
BSM-Spanish	-.23	.35	*.60*
WPPSI Block Design	.13	*.78*	.04
Verbal Fluency	-.09	*.63*	.35
Verbal Memory-1	.01	*.63*	.42
Verbal Memory-3	-.45	*.46*	.40
Arm Coordination	*.89*	.17	.03
Draw-A-Child	.05	*.75*	-.01
PIPS	-.14	.11	*.75*
POCL	.28	.01	*.79*

[a]Principal components solution, varimax rotation.

Comparisons with previous factor analyses. The results of the current analyses are not directly comparable to the spring 1976 analyses, since Child Rating Scale data (for the English-dominant sample) were included at that time. However, there are interesting similarities. Arm Coordination maintains its position as a distinct dimension, and language tasks and the social problem-solving measures (PIPS, POCL) are clustered in a similar manner. The resemblance between the factor structures found for the two language samples recommends the possibility of equating English and Spanish versions of the batteries in the future.

Characteristics of the Classroom Observation System

The PDC Classroom Observation System differs in many ways from the other instruments in the battery of child measures: it is not a test of performance under special conditions but rather a record of performance under natural conditions; it is conceived as reflecting characteristics of the classroom environment as much as it reflects characteristics of the children who comprise the class; its scoring and the methods by which its reliability and validity are established are unique to this measure among those in the battery. Thus in this section the Observation System is discussed separately from the other instruments.

The PDC Observation System was developed to provide descriptive information regarding the social-emotional competence of children in their classroom settings. The behavior categories that make up the instrument were formed by redefining, and in some cases, combining, behavior categories from existing observation instruments that differentiate between children of varying degrees of social competence, and by adding other categories appropriate to PDC goals. The theoretical rationale for selecting these categories is that they measure a "general attitude of negotiation and reciprocity in dealing with others in a social environment."[1] This attitude is believed to be generalizable across all cultural groups and implies that a child's own needs and goals are valuable, but that the needs and goals of others are equally important and must be taken into account. More specifically, the developing child should learn how to control and influence others with effective strategies that do not violate the rights of others. (For example, physical force is

---

[1] Bronson, M. Executive competence in preschool children. Paper presented at the meeting of the American Educational Research Association, Washington, D.C., 1975. For a more extensive listing of references to the literature consulted in developing the system, see Interim Report II, Part B (June 1975).

considered to be a violation of others, and thus does not indicate an attitude of negotiation and reciprocity.) In addition, the child should be reasonably influenced by others, but not totally subservient to or dominated by them. Other social strategies that promote and sustain social interaction such as sharing, helping, requesting and providing resources, and taking turns are also considered important indications of a child's social competence and are represented in the categories of the observation system. Definitions and examples of the observation categories are presented in Appendix F.

Summary of Instrument Development

Fall 1975 and spring 1976 observation data collection efforts were aimed at establishing the psychometric properties of the observation instrument. After each collection, observers reported that the instrument could be used in the field with little difficulty. In addition, spring reliability assessments have shown that observers can be trained to use the instrument with a desirable level of accuracy. This implies that observation categories have been sufficiently defined and clarified that minimal inference is required by the observer when coding specific behaviors.

Analyses of fall 1975 and spring 1976 observation data have shown some degree of relationship between children's social-emotional, psychomotor, and cognitive competences, i.e., children's observed social skills corresponded to their performance on other measures in the PDC battery. But in view of the low magnitude of these relationships, particularly with regard to teachers' ratings of similar dimensions of children's behaviors, the validity of the instrument as a measure of children's social skills could not be adequately established.

In part, this absence of validity may be attributed to the way in which behaviors were sampled. Since the activity level of the classroom (i.e., opportunity for social interactions) was found to be highly related to children's behaviors, it seems likely that the 20-minute observation period provided a description of the child's behavior under only one particular classroom condition. Under other conditions, the child's behavior might have been very different from that sampled. The time-sampling technique, then, could account for the low correspondence between a teacher's assessment of a child's social skills and the description provided by the observation instrument.

For this reason it was proposed in Interim Report IV (August 1976) that the Observation System can be regarded as a measure of classroom "personalities," rather than of individual behaviors, and that it be used for assessing PDC impact at a classroom level of analysis.

Purpose of this Analysis

This analysis is directed toward establishing the psychometric properties of the Observation System as a measure of classroom characteristics. One step in this process is to substantiate the findings that observers can be trained to use the instrument with acceptable levels of accuracy. Thus, as in the past analyses, reliability data were collected and analyzed for this report. A second step is to assess whether the instrument accurately represents, and thus adequately measures, the classroom. This was determined by pairing observation subcategories with other measures in the PDC battery and assessing their relationships using classroom-level analyses.

It is also important to establish the value of the instrument for measuring PDC's impact on children's classroom behaviors or on classroom conditions. Exploratory analyses were thus conducted which examined the comparability of PDC and comparison classrooms in fall 1976. No consistent differences were expected, since this is the first operational year of PDC, but these analyses were intended to provide a check on the initial equivalence of PDC and comparison classrooms on such dimensions as frequency of social interaction, amount of time spent in nonsocial activities, amount of verbal behavior, and opportunity for social interactions. These comparisons also provided a test of the analytic methods available for investigating classroom comparability.

Observation Procedures

In an attempt to insure that observation data would be collected in a consistent manner across sites, guidelines and procedures for completing observations were specified in detail during the observer training session.

Before they began their observations in the classrooms, observers met with classroom teachers to describe the observation instrument and answer questions. To control for any observation bias, the observers completed all observations prior to administering the child tests and observed only the children who were listed on their rosters.

Beginning with the first child on their rosters, observers observed each child for two consecutive five-minute intervals. Each five-minute interval was divided into fifteen 20-second units. These units were further divided: 5 seconds for observing and 15 seconds for recording. The

observing and recording intervals were signalled by a portable cassette tape recorder that emitted an electronic "beep" into an earphone worn by the observer.

The number of observations completed per day varied with the number of children in the classroom, the class schedule for that particular day, children's absences, and activities that took the children outside the classroom. Observers were advised to observe during all periods of the day except outdoor play and toileting. If, however, regular classroom activities such as storytime, art, or snacktime were conducted outdoors, they were instructed to observe during those times as well. Observers also received instructions on how to handle situations that might interrupt their observations and on how to handle child absences.

Fall 1976 Observation Training Procedures

The primary objective of the September 1976 training session was to adequately train observers, especially new ones, so that reasonable coding reliability could be achieved. To the extent possible, only those observers who had received training and collected observation data in fall 1975 and/or spring 1976 were included in this training session.[1] This allowed new training methods to build upon the group's previous training and experience in the classroom.

As in other observation training sessions, the Observation System was introduced in a large-group session during which changes and revisions were noted and examples of observation categories were provided. Small-group sessions were scheduled to explain and give examples of the observation categories. Throughout these sessions, observers were asked to describe and role-play examples of behaviors they had observed in previous classroom observations; trainers then indicated how the behaviors should be coded on the record sheet. Additional small-group sessions were used for viewing videotapes of preschool-aged children in school settings. After the observers had coded a two-minute segment of the tape, trainers provided feedback on how the behaviors should have been coded. Common errors made by the participants were discussed and additional clarification and examples were provided for ambiguous or frequently confused categories.

---

[1] Only five of the observers who were selected to collect fall 1976 observation data had neither attended other observation training sessions nor collected observation data previously.

Reliability of the Observation System

Collection of reliability data. The reliability estimates included in this report were gathered at the end of the fall 1976 training session. This reliability assessment was necessary for determining how well observers were prepared to begin their observations in the field. To assess this, observers simultaneously watched and independently coded a 40-minute videotape which included several clear examples of the behavioral categories contained in the observation system.

Analysis of reliability data. The accuracy of observation coding was assessed by comparing each observer's responses to a criterion coding of the same behavioral events. Although the measure produced is not identical to a conventional measure of inter-observer agreement, it does assess the accuracy of observers' coding compared to a single standard criterion. This provides a basis for detecting those categories that were commonly coded unreliably by a majority of the observers. Further, analytical inferences that include these categories could take this factor into account.

Two methods for assessing coding reliability were employed. The first method computed a pairwise observer and criterion agreement estimate within categories for each 5-second observation interval. A proportion of agreement was determined using Cartwright's alpha.[1] This procedure consists of comparing, unit by unit, the codes selected by the observer with the criterion codes. Estimates were obtained for the number of times observer and criterion codes agreed and disagreed for each observation unit. The reliability was then computed by dividing the number of times codes agreed by the number of agreements plus disagreements.

The second method of reliability assessment examined how well observers' codes matched the criterion for the total length of a given observation period. (Although the observers' codes may not agree with the criterion unit by unit, it is important that observers, after viewing a child for a specified interval, at least agree on the relative number of tallies assigned to each subcategory.) To obtain this estimate, the mean total number of tallies assigned to a given category was divided by the criterion number for that category (i.e., the number of times behavior of that type was actually exhibited in the reliability tape). Different inferences can be drawn from these two reliability estimates: the first, based upon Cartwright's alpha, indicates the reliability of a single observation within a specified category; the second, overall proportion of agreement, indicates the reliability of the total frequency of observations for a category.

---

[1] Cartwright, D. S. A rapid non-parametric estimate of multi-judge reliability. Psychometrika, 1956, 21, 17-29.

Reliability results. Table 14 presents the mean proportion of agreement and Cartwright's alpha for each subcategory of the observation system. Also included in this table is an indication of whether observers overestimated or underestimated the frequency of a specific subcategory.

As has been found in past analyses, it appears that observers can accurately distinguish and code children's involvement in the classroom (Noninvolved, Social, and Nonsocial). Proportions of agreement for these subcategories ranged from .96 to .99, with a mean of .98. These reliability estimates are higher than those reported for these subcategories in spring 1976. The high agreement figures, .99, for verbal behavior (Verbal English and Nonverbal) found in the current analysis are also noteworthy.

While there is a slight increase in reliability estimates from spring to fall, errors continue to occur most frequently in those subcategories describing the child's behavior during child-peer and child-adult interactions. Proportions of agreement in these subcategories ranged from .57 to .98, with a mean of .81. However, considering the five distinctions that these categories require observers to make concerning the nature and purpose of the child's social interactions, some imprecision is expected. Thus the coding accuracy figures seem acceptable.

Items that describe classroom activity (Maximal, Moderate, and Minimal) were excluded from this reliability assessment because of difficulties in portraying the general activity level of a classroom on videotape. However, last spring's onsite reliability assessment indicated that observers could accurately assess and code a child's opportunities for social interaction in the classroom. For that assessment, the mean reliability estimate for this category was .88. Because this category was not revised, it is assumed that this coding accuracy was maintained for fall data collection.

Overall, there was a substantial improvement in observers' observation and coding skills. In comparing fall and spring reliability estimates, the accuracy of coding increased or remained the same for 80% of the observation categories.

Analysis of Classroom Observation Data

Preparation of observation data for descriptive analysis. For all observation categories, a sum of the child's behavioral incidents across the 10-minute observation intervals was computed. Each child, then, had one summary score for every item on the observation instrument. These scores were then summed across

Table 14

Coding Reliability
Fall 1976

Observation Category	Number of Examples	Proportion of Agreement	Cartwright's Alpha	Direction of Error[a]
Noninvolved	6	.98	.59	+
Social	110	.99	.97	+
Nonsocial	13	.96	.68	-
Verbal English	84	.99	.96	+
Verbal Spanish	0			
Verbal Combined	0			
Nonverbal	39	.99	.88	-
Negative	9	.67	.66	-
Positive Control	22	.88	.71	+
Positive Resist	8	.98	.75	+
Other Positive Behaviors	26	.97	.68	+
Requests Information	6	.98	.76	-
Gives Information	10	.65	.54	+
Requests Assistance	22	.66	.66	-
Gives Assistance	10	.89	.50	+
Requests Support	7	.72	.73	-
Gives Support	3	.87	.86	-
Other Purposes	7	.71	.50	+
Negative	5	.71	.79	-
Positive Control	22	.97	.79	-
Positive Resist	8	.71	.70	-
Other Positive Behaviors	15	.82	.64	+
Requests Information	7	.96	.81	-
Gives Information	9	.84	.68	+
Requests Assistance	11	.84	.72	-
Gives Assistance	9	.95	.68	+
Requests Support	5	.57	.63	-
Gives Support	1	.67	.57	-
Other Purposes	8	.88	.60	+

[a]Overestimations and underestimations of individual categories are represented respectively by plus and minus signs.

children within classrooms, creating a classroom mean for each category. Because past observation analyses indicated that children's behaviors varied according to the activity level of the classroom, a classroom score should take this into account. Therefore, as in past analyses, relative frequencies were computed for each activity level, weighting the absolute scores by the amount of time classrooms were observed in a given activity level. The means and standard deviations of these transformed variables are given in Appendix F.

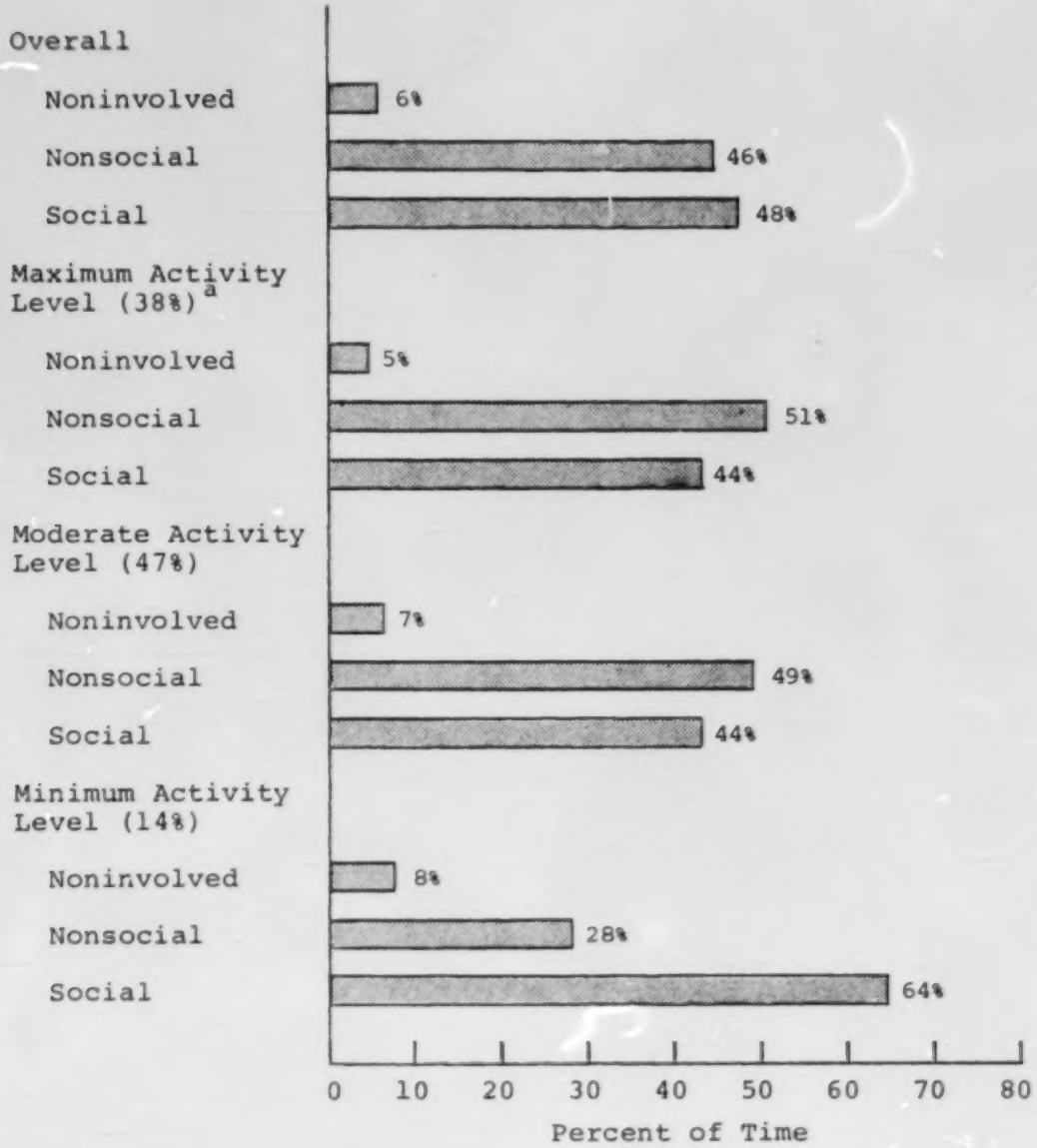
Results of descriptive analyses. In order to provide a summary description of the observation data, the classroom relative frequency for each observation variable was computed for each activity level. Classroom frequencies were then aggregated across sites. Figure 8 presents the relative amount of time children in these classrooms were engaged in Noninvolved, Social, and Nonsocial activities. The results indicate that these groups spent high proportions of their time in activities that involved objects (46%), and in activities that involved persons (48%). These proportions were found to vary across activity levels with Minimal activity levels proving most conducive to social interactions. Interactions during Minimal activity periods, however, often occurred in a large-group setting where children were attending to the adult leading the group activity. These social interactions differ qualitatively from those that would occur during free-play or small-group time (i.e., Maximal and Moderate classroom activity levels).

Figure 9 shows the relative frequencies of children's verbal behavior for those classrooms in which the majority of children spoke English. Again, as in past analyses, children were verbal less than 30% of the time, with some variation across activity levels. Out of the total time children were verbal, 52% of this verbal activity occurred during peer interactions, while 48% occurred during adult interactions.

Figure 10 shows the relative frequencies of children's verbal behavior for Spanish-speaking classrooms. For these classrooms, children were verbal less than 40% of the time, with only slight variations across activity levels; Spanish was spoken 21% of the time and English was spoken 17% of the time. Out of the total time Spanish was used in social interactions, 66% of this verbal behavior occurred during peer interactions, while 34% occurred during adult interactions. There was no difference in the relative amount of time children spoke English (20%) or Spanish (20%) during adult interactions; however, during peer interactions Spanish was spoken more frequently than English (33% vs. 27%).

Figure 8

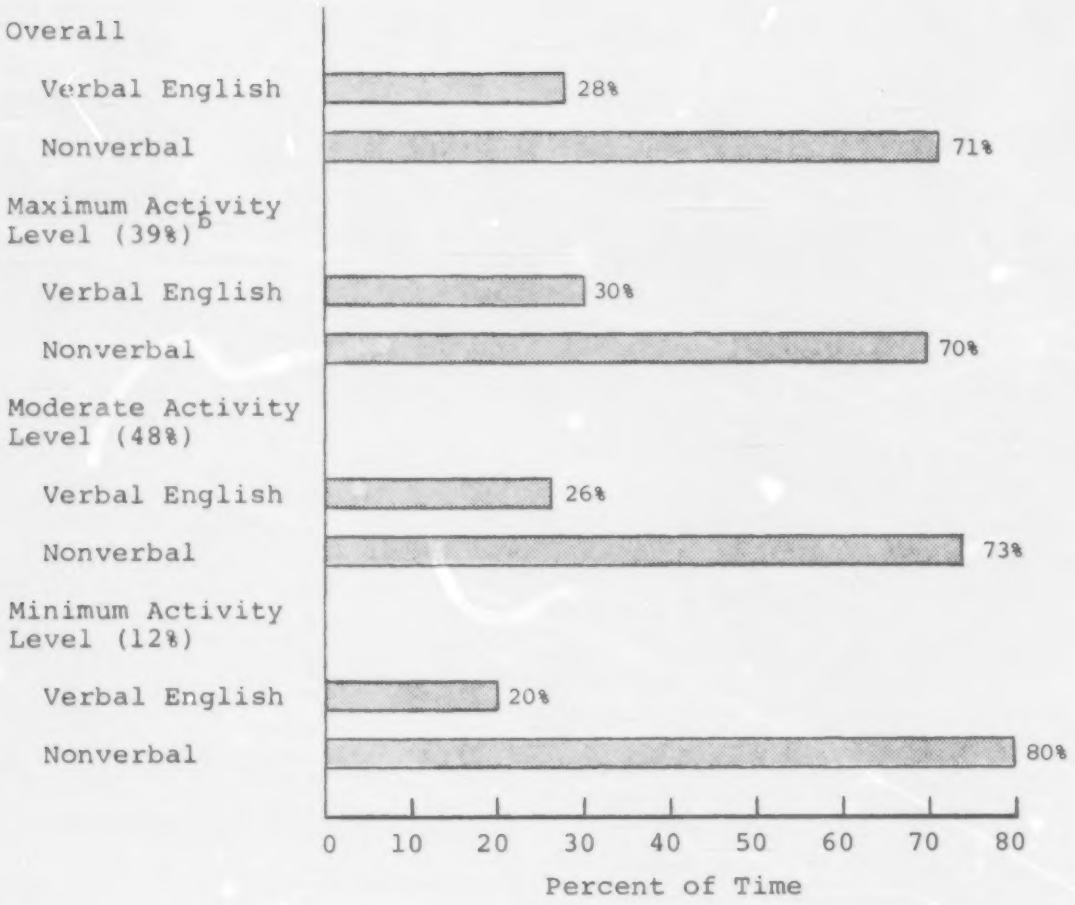
Relative Frequencies of Classroom Involvement
by Classroom Activity Level
Fall 1976
N = 80



[a]Percentage indicates the relative amount of time classrooms were observed under conditions that respectively permit maximum, moderate, and minimum opportunity for social interactions.

Figure 9

Classroom Verbal Behavior: English-Speaking Classrooms[a]
Fall 1976
N = 71

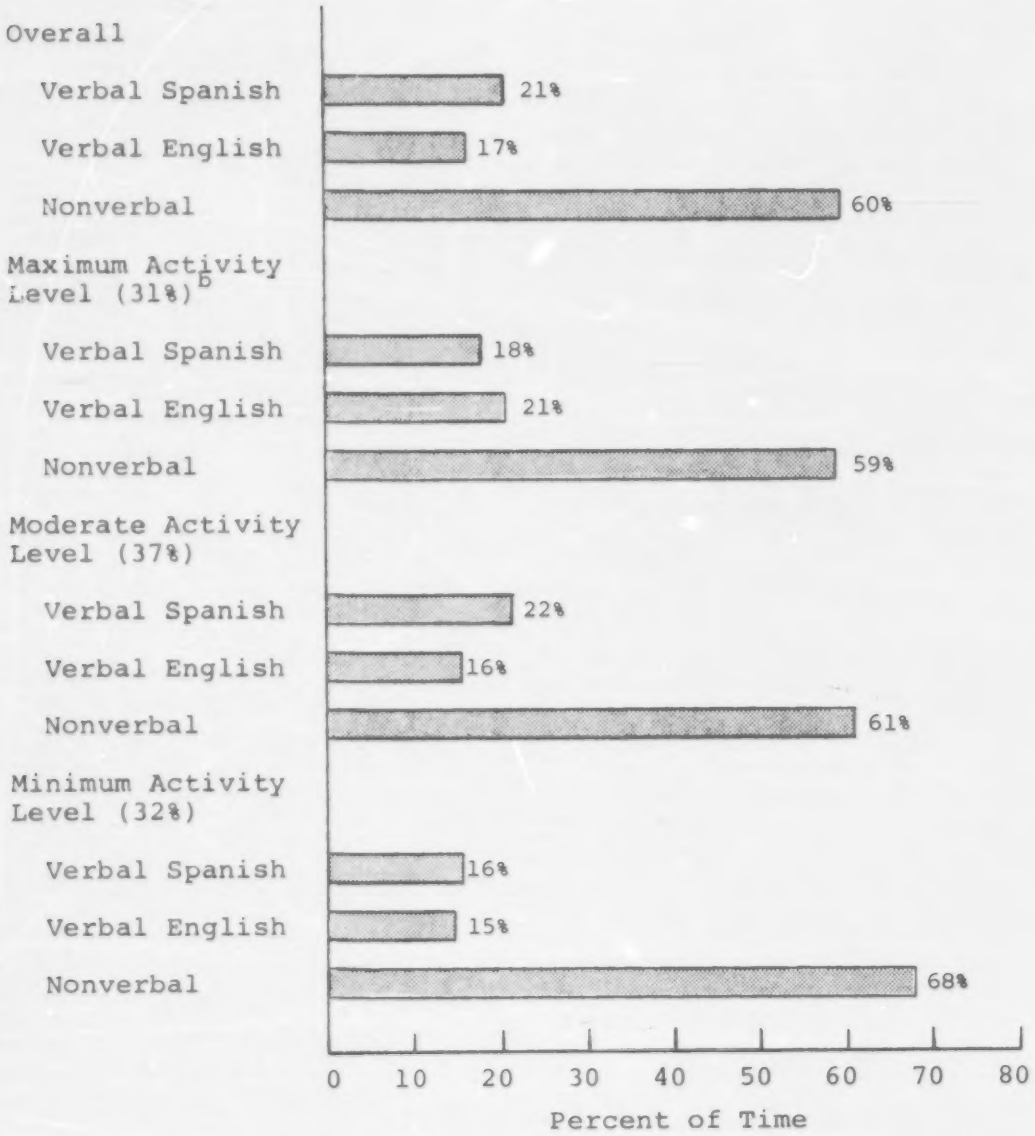


[a]These are classrooms in which the majority of children spoke English.

[b]Percentage indicates the relative amount of time classrooms were observed under conditions that respectively permit maximum, moderate, and minimum opportunity for social interactions.

Figure 10

Classroom Verbal Behavior: Spanish-Speaking Classrooms[a]
Fall 1976
N = 9



[a]These are classrooms in which the majority of children spoke Spanish.

[b]Percentage indicates the relative amount of time classrooms were observed under conditions that respectively permit maximum, moderate, and minimum opportunity for social interactions.

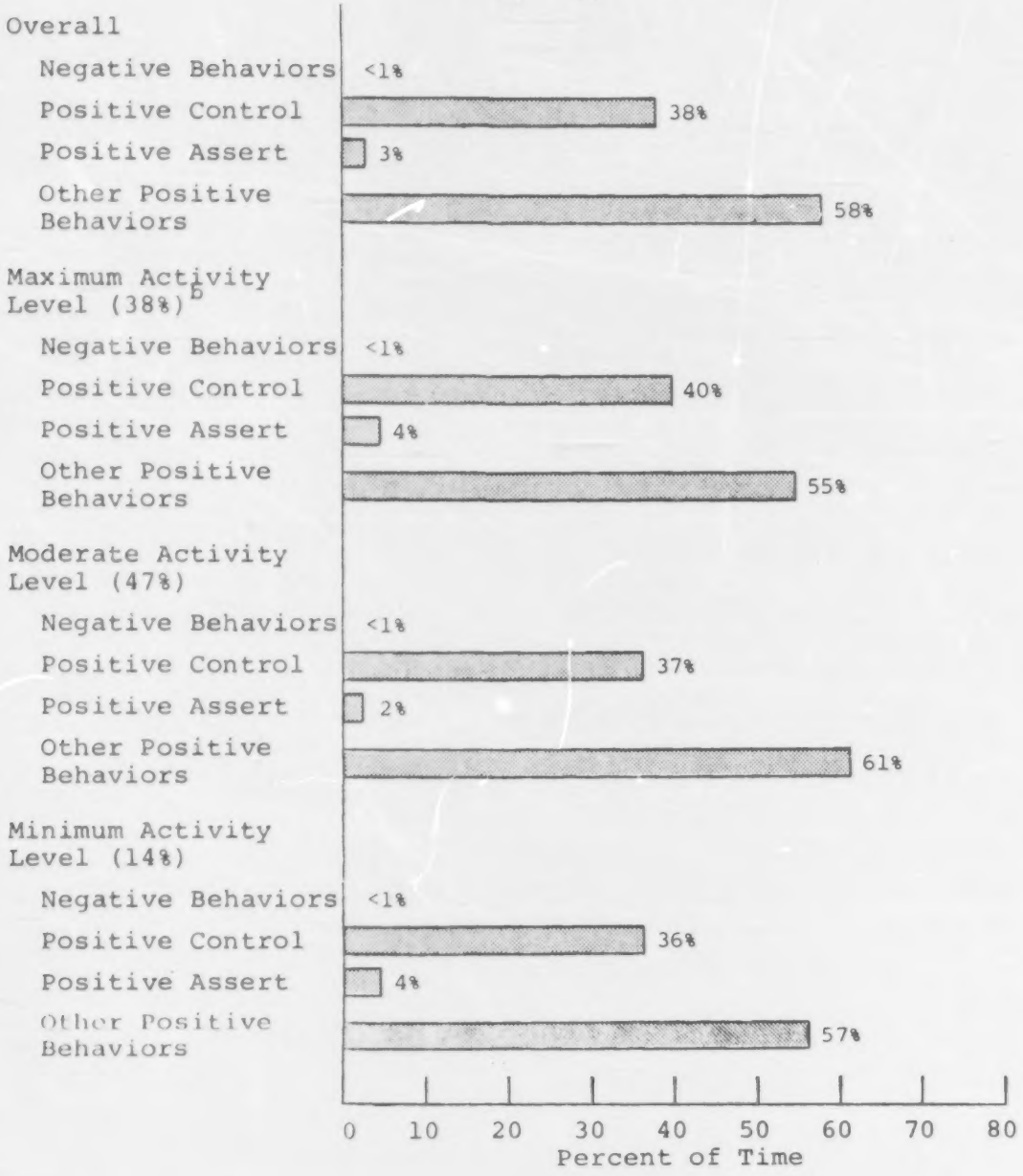
Figures 11 and 12 display the relative frequencies of child-peer and child-adult classroom interactions. Classroom social interactions more often involved adults (50%). Peer interactions occurred 45% of the time, while joint interactions with both a peer and adult occurred 5% of the time. As in past analyses, it appears that the majority of children's social interactions with peers and adults were positive in nature, as negative behaviors were exhibited less than 1% of the time. The rates of children's positive attempts to Control peers (33%) or adults (34%) are higher than any previous analysis has shown. Adult Controls fluctuated substantially across activity levels; peer Controls were only slightly influenced by the activity level of the classroom. Further, the relatively large standard deviation (.20) of these proportions suggests that there is some variability across classrooms in the frequency of controlling incidents.

Classroom behaviors reflecting children's positive attempts to resist the control of others (i.e., Assert) occurred infrequently: 2 to 4% of the time. Because the occurrence of this behavior is associated with the amount of control exhibited by others, the low frequency of Asserts behaviors, implies either that children were not directed by others (which is contrary to the evidence of increased controlling in the classroom), or that they have not learned acceptable ways of resisting this control. Thus, it appears they are resisting control in a negative manner (coded Negative) or simply complying with others' directives (coded Positive). However, last spring's observation data reflected a higher frequency of assertive behaviors, which suggests that the new sample of children may learn these strategies as they gain more experience in Head Start classrooms.

Figure 13 describes child-adult and child-peer interactions from another perspective: purpose of interaction rather than nature of interaction. High proportions of Gives behaviors were exhibited. During adult interactions, children provided information 45 to 59% of the time, and provided assistance or materials 6 to 8% of the time. A high incidence of Gives Information (48% to 51%) and Gives Assistance (7 to 10%) was also found for peer interactions. In contrast, Requests behaviors were exhibited in the classroom less than 30% of the time. These behaviors were primarily directed toward obtaining assistance or materials from either an adult (13%) or peer (17%); requesting or providing emotional support occurred less than 1% of the time for both adult and peer interactions. Classroom patterns of Gives and Requests varied only moderately across activity levels. There was, however, some evidence that as activity levels changed from Maximal to Minimal conditions for social interactions, Gives Information (within the context of adult interactions) increased. Because instructional activities (e.g., music, storytime) typify Moderate and Minimal conditions, it is not surprising that the incidence of providing information increased for these activity levels.

Figure 11

Relative Frequencies of Classroom Child-Peer Interactions: Nature of Interaction by Classroom Activity Level[a]
Fall 1976
N = 80

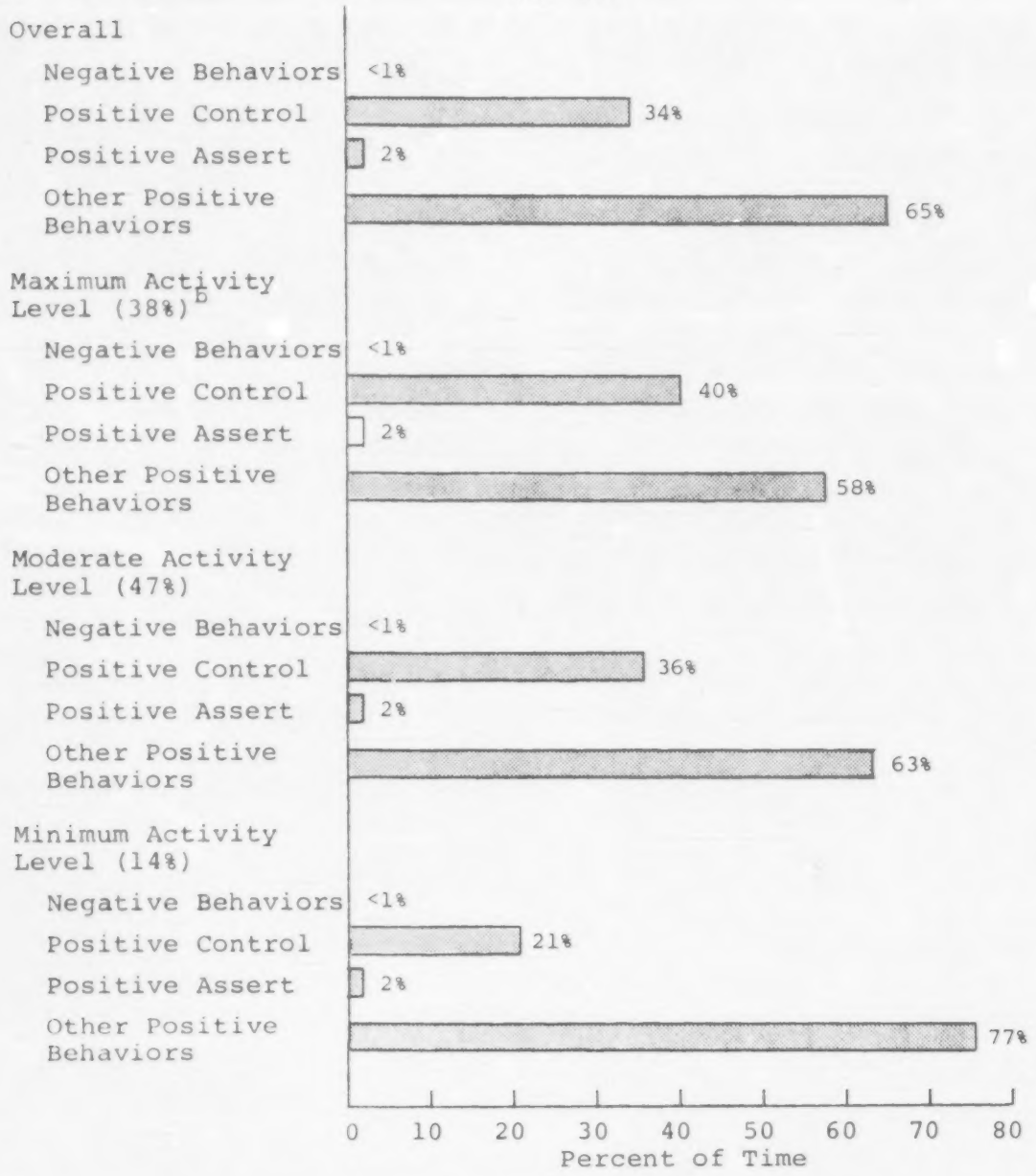


[a]Child-Peer interactions represent 45% of classroom social activities.

[b]Percentage indicates the relative amount of time classrooms were observed under conditions that respectively permit maximum, moderate, and minimum opportunity for social interactions.

Figure 12

Relative Frequencies of Classroom Child-Adult Interactions: Nature of Interaction by Classroom Activity Level[a]
Fall 1976
N = 80



[a]Child-Adult interactions represent 50% of classroom social activities.

[b]Percentage indicates the relative amount of time classrooms were observed under conditions that respectively permit maximum, moderate, and minimum opportunity for social interactions.

Results of correlational analyses. The correlations shown in Table 15 were computed to assess the interrelationships among observation variables at the classroom level.[1] For this analysis, particular attention was given to determining whether findings from past analyses hold for classroom-level data or whether new relationships emerge.

As in past analyses, children's rates of verbal behavior and the strategies they use while interacting with others were only slightly related. This finding suggests that, for this young age group, children control others and resist the control of others in a nonverbal fashion. Also, children often request assistance from others nonverbally, as a significant correlation was found between the Nonverbal and Requests Assistance categories. Other significant correlations were found between children's controlling behaviors and the frequency of Requests Assistance. Thus, as has been found in past analyses, children's controlling behaviors were generally directed toward obtaining help or materials from an adult or peer; that is, they were directed toward some end or goal.

A departure from past findings, though, is the significant intercorrelations between peer and adult interaction items were found in this analysis. In classrooms where there was a high frequency of peer Controls, there was also a high frequency of adult Controls. A similar relationship was found between peer and adult requests for assistance.

Of additional interest are the relationships between observation variables and other child measures. To examine these, classroom means were computed for the child tests and correlated with observation variables.[2] As shown in Table 16, the relationships found ranged from moderately negative to moderately positive. Observation categories that occurred more frequently (e.g., Social, Nonsocial, Verbal English, Nonverbal) tended to show higher correlations with other variables.

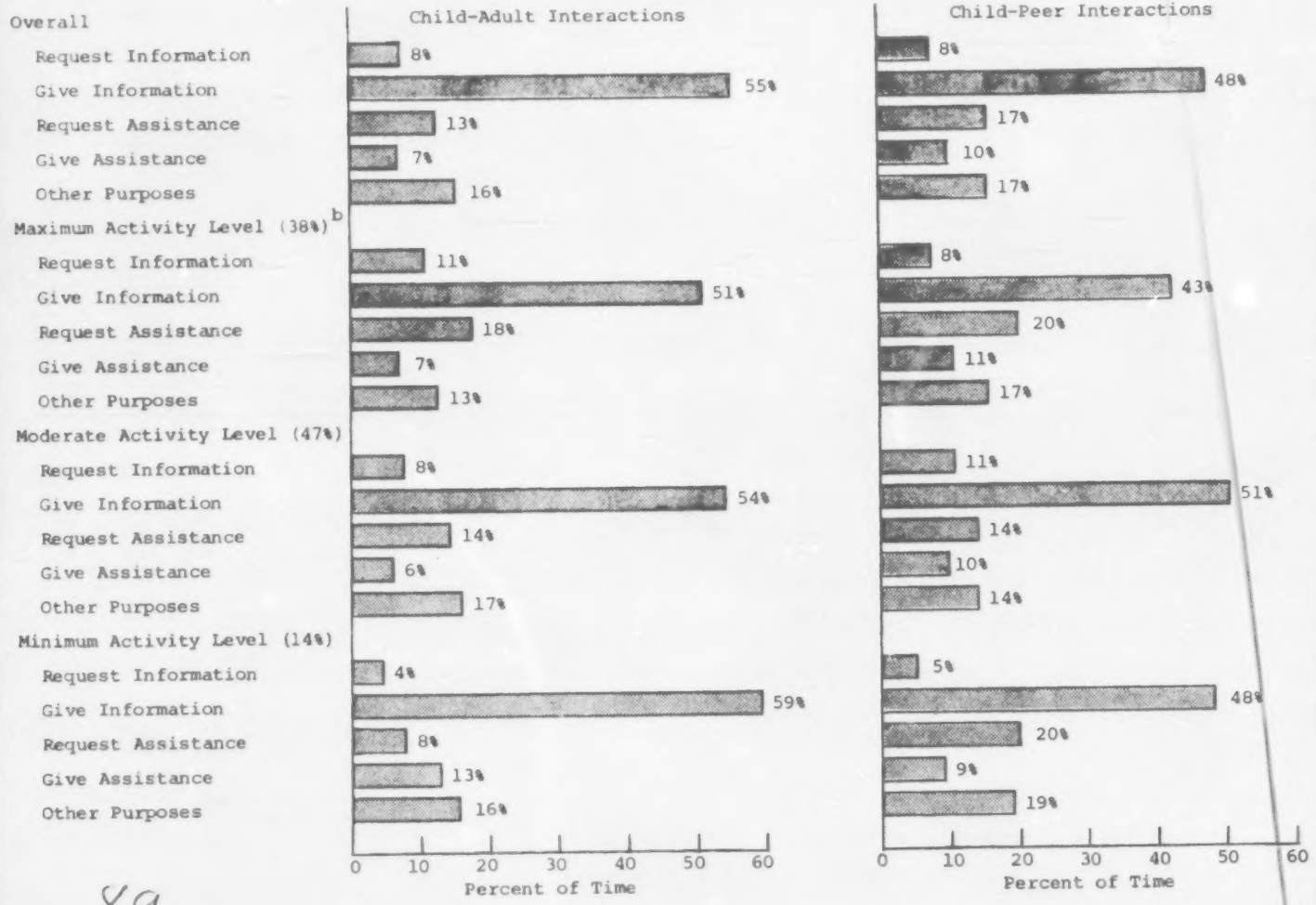
---

[1]For this analysis, children's scores on the observation variables were averaged within classrooms. The classroom then became the unit of analysis.

[2]Only classrooms in which the majority of children spoke English were used in this analysis.

Figure 13

Relative Frequencies of Classroom Adult and Peer Interaction: Purpose of Interaction by Classroom Activity Level[a]
Fall 1976
N = 80



89

[a]Child-Adult interactions represent 50% of classroom social interactions, while Child-Peer interactions represent 45% of classroom social interactions.

[b]Percentage indicates the relative amount of time classrooms were observed under conditions that respectively permit maximum, moderate, and minimum opportunity for social interactions.

Table 15

Intercorrelations of Observation Variables
Fall 1976

N=80

		Classroom Involvement			Verbal Behavior			Description of Peer Interactions					Purpose of Peer Interaction				Description of Adult Interaction					Purpose of Adult Interaction				Classroom Activity Level	
		Noninvolved	Social	Nonsocial	English	Spanish	Nonverbal	Rate of Interaction	Positive Control	Positive Assert	Other Positive Behaviors	Negative Behaviors	Requests Information	Gives Information	Requests Assistance	Gives Assistance	Rate of Interaction	Positive Control	Positive Assert	Other Positive Behaviors	Negative Behaviors	Requests Information	Gives Information	Requests Assistance	Gives Assistance	Maximum	Moderate
Classroom Involvement	Noninvolved																										
	Social	-.29																									
	Nonsocial	-.18	-.89																								
Verbal Behavior	English	.03	-.02	.00																							
	Spanish	-.17	.42	-.35	-.48																						
	Nonverbal	.15	-.43	.37	-.47	-.55																					
Description of Peer Interactions	Rate of Interaction	.00	-.19	.19	-.02	-.16	.19																				
	Positive Control	-.33	.39	-.24	-.12	.18	-.10	.01																			
	Positive Assert	-.14	-.13	.20	-.03	.09	-.05	-.08	-.03																		
	Other Positive Behaviors	-.33	-.37	.21	.11	-.19	.19	.00	-.98	-.15																	
	Negative Behaviors	.12	.01	-.07	.17	.03	-.18	.05	-.08	.13	-.01																
Purpose of Peer Interaction	Requests Information	.23	-.14	.03	.25	-.12	-.11	.11	-.27	.01	.27	.05															
	Gives Information	.16	.24	-.33	.18	-.24	.06	-.33	-.01	-.22	.06	-.28	-.08														
	Requests Assistance	-.25	.08	.04	-.19	-.04	.18	.34	.35	.07	-.36	.09	-.09	-.36													
	Gives Assistance	-.11	-.16	.23	.00	.21	-.20	.06	-.01	.16	-.03	.21	-.25	-.47	-.13												
Description of Adult Interaction	Rate of Interaction	.09	-.06	.02	.12	.08	-.19	-.93	-.11	.14	.08	.01	-.06	.26	-.39	.03											
	Positive Control	-.25	.14	-.02	-.18	.13	.05	.08	.31	-.13	-.27	-.14	-.23	.01	.00	-.16	-.13										
	Positive Assert	-.12	.01	.05	.14	-.05	-.09	.12	.05	.09	-.08	.19	-.09	.03	.04	.04	-.15	-.13									
	Other Positive Behaviors	.28	-.14	.01	.14	-.12	-.03	-.11	-.32	.11	.29	.09	.25	-.02	-.01	.15	.17	-.97	-.10								
	Negative Behaviors	.10	.01	-.06	.12	-.04	-.08	-.03	-.09	.09	.08	-.05	.00	.20	-.09	-.10	-.02	-.07	.08	.04							
Purpose of Adult Interaction	Requests Information	.08	-.04	.01	.16	.06	-.19	-.04	-.32	-.10	.34	-.03	.26	.05	-.17	-.16	.05	.09	-.15	-.06	-.01						
	Gives Information	.06	.18	-.22	.11	-.19	.06	-.24	.20	-.08	-.18	-.15	.13	.48	.00	-.38	.20	-.13	.07	.11	.13	-.17					
	Requests Assistance	-.16	-.11	.19	-.20	-.04	.23	.22	.01	.03	-.01	-.06	-.21	-.26	.25	.06	-.22	.52	-.13	-.49	-.09	-.12	-.41				
	Gives Assistance	.20	-.09	-.01	.02	.26	-.27	.20	-.16	.10	.12	.26	-.11	-.40	-.18	.60	-.12	-.18	.10	.16	.01	-.08	-.48	-.05			
Classroom Activity Level	Maximum	-.31	-.31	.47	-.03	-.09	.14	.26	-.06	.01	.07	-.10	-.14	-.20	.12	.11	-.25	.21	.18	-.25	-.10	.25	-.29	.23	-.06		
	Moderate	-.04	.09	-.08	.06	.00	-.06	-.30	.01	.22	-.04	-.13	.18	.18	-.16	-.15	.32	-.21	-.15	.24	.08	-.19	.29	-.18	-.19	-.69	
	Minimum	.45	.27	-.50	-.04	.12	-.11	.04	.06	-.30	-.03	.28	-.06	.02	.06	.04	-.09	.00	-.03	.01	.02	-.06	.00	-.07	.32	-.38	-.40

NOTE: Correlations higher than +.22 or lower than -.22 are statistically significant ($p<.05$).

Table 16
Correlation of Observation Variables and Other Reliable Measures[a]
Spring 1976
(N=70)

	Classroom Involvement			Verbal Behavior		Descriptors of Peer Interactions					Purpose of Peer Interactions				Descriptors of Adults Interactions					Purpose of Adult Interactions			
	Noninvolved	Social	Nonsocial	English	Nonverbal	Frequency of Interaction	Positive Control	Positive Assert	Other Positive Behaviors	Negative Behaviors	Requests Information	Gives Information	Requests Assistance	Gives Assistance	Frequency of Interaction	Positive Control	Positive Assert	Other Positive Behaviors	Negative Behaviors	Requests Information	Gives Information	Requests Assistance	Gives Assistance
Arm Coordination	00	24*	-24*	06	-09	06	24*	-02	22	-06	07	-01	16	06	-06	03	-03	-02	-17	-32*	28*	-12	13
BSM-English	11	-07	01	24*	-25*	-23	00	-01	01	-02	-04	02	-47*	16	29*	-11	-14	14	-03	17	00	-27*	21
Draw-A-Child	-01	26*	-24*	18	-28*	-33*	05	07	-05	-27*	01	12	-28*	01	33*	-13	-24*	20	-02	09	17	-34*	13
PIPS	32*	00	-17	03	00	-08	-29*	-21	32*	01	20	10	-07	-09	04	-21	-30*	29*	04	20	-04	-10	-08
POCL-1	22	-38*	25*	-31*	32	03	-09	-01	14	-21	27*	-09	-01	-09	-09	04	-16	01	-10	14	-20	00	06
POCL-2	11	-26*	20	-26*	23*	07	-02	-10	05	-18	12	-09	01	-10	03	07	-15	-04	-05	06	-26*	13	11
POCL-Total	20	-36*	24*	-31*	31*	01	07	-19	12	-21	24*	-09	00	-10	-06	05	-17	-01	-09	12	-22	04	08
Verbal Fluency	16	11	-19	12	-13	-07	-11	-16	16	-29*	-04	24*	-11	04	05	-20	-13	23*	-06	05	-07	-12	21
Verbal Memory-1	21	02	-14	19	-16	-13	-16	-22	22	-35*	-06	27*	-23*	00	14	-27*	-09	29*	-05	-03	10	-21	21
Verbal Memory-3	12	03	-09	05	-05	-07	-16	-09	18	-11	-08	26*	-19	07	10	-18	-08	19	04	24*	00	-10	02
WPPSI Block Design	20	11	-21	12	-18	-20	-21	-09	24*	-22	01	16	-34*	06	19	-10	-20	14	19	17	17	-29*	25*

*p < .05 NOTE: For graphic convenience decimal points have been omitted from correlation coefficients.

[a]Only classrooms in which the majority of children spoke English were used in this analysis.

Although behavior in the individual subcategories describing peer and adult interactions in the classroom occurred less frequently, some of these categories were moderately, and in some cases significantly, related to other child measures. In particular, rates of peer and adult positive behaviors were positively related to scores on the PIPS. Additionally, rates of requesting help from adults or peers were significantly but negatively correlated with BSM, Draw-A-Child, and WPPSI scores. It appears that the rate of requests for assistance is inversely related to a wide range of children's cognitive and language competencies.

The findings from this analysis indicate that a number of classroom-level interaction patterns are related to performance on psychomotor, cognitive, and language measures. Inasmuch as higher degrees of association were found for those categories describing global characteristics of the classroom (e.g., Involvement, Verbal Behavior, Rate of Adult Interactions), greater confidence can be placed in the findings related to these categories.

Results of comparability analyses. The assumption made in identifying certain Head Start centers and schools as comparison institutions is that they are essentially the same to begin with as their PDC counterparts. Therefore, since last fall marked the start of PDC's first operational year--its official beginning--the classroom observations made in the fall were expected to indicate that no systematic differences exist between PDC and comparison classrooms. To confirm (or disconfirm) this expectation, analyses were performed that examined the measures obtained for PDC and comparison classrooms in ten categories with the highest incidence of occurrence. These were rate of:

Noninvolvement,
Social Involvement,
Nonsocial Involvement,
Verbal English,
Nonverbal Behavior,
Peer Interactions,
Adult Interactions,
Maximal, Moderate and Minimal Classroom Activity.

Two alternate types of analyses,[1] parametric and non-parametric, were conducted; the results of one provided cross-checks on the results of the other. Both procedures were consistent in showing no significant differences between groups, with one exception:

---

[1]The analytic procedures that led to the results described here are discussed in Appendix H.

the non-parametric procedure indicated that children in comparison classes spend a significantly greater proportion of their time in nonverbal activities (the parametric procedure produced a finding of no significant difference on this variable). Considering the fact that a total of 20 significance tests were performed (10 variables x 2 statistical procedures), a finding of only one significant difference indicates that PDC and comparison classes are indeed very much the same on the dimensions measured by the Observation System.

IV

CONCLUSIONS

Adequacy of the Tests and the Samples

The major purposes of this report are, as noted, to answer the questions of instrument appropriateness, PDC-comparison group comparability, and sample size adequacy. The answers to be offered here are tentative in two ways: they anticipate the conclusions that will emerge from the meeting of the PDC Evaluation Advisory Panel in April, and they anticipate the conclusions that will be drawn by OCD, which bears the ultimate responsibility for decisions made about PDC.

Are the Measuring Instruments Appropriate to the Task?

Child measures, individually. It can be said with few reservations that all of the instruments included in the battery satisfy all the criteria that have been used in judging them. As Table 17 reflects, the reservations are minor: Verbal Memory-1 and Draw-A-Child were earlier found inappropriate for older children but, with changes in the content of the former and the scoring of the latter, it is believed that they will now be appropriate; the English version of the Bilingual Syntax Measure may not be adequately sensitive to change for children whose dominant language is Spanish, and the Spanish version may not be adequately sensitive for English-dominant children.

Child measures, collectively. Although the factor structure of the battery does not correspond exactly to the *a priori* categorization of the tests (cognitive-language, social-emotional, psychomotor), the factors that emerge are similar to those expected for both the English-dominant and Spanish-dominant groups, and indicate that the battery does provide coverage of these areas.

Table 17

Summary of Findings on Characteristics of Tests Included in the Fall 1976 Battery

	Internal Consistency	Validity	Sensitivity to Change[a]	Relevance to Social Competence[a]	Developmental Span[a]	Ease of Administration
BSM-English	✓	✓	(✓)	✓	-	✓
BSM-Spanish	✓	✓	(✓)	-	-	✓
Block Design (WPPSI)	✓	✓	-	-	-	✓
Verbal Fluency	✓	✓	✓	✓	✓	✓
Verbal Memory-1	✓	✓	✓	✓	(✓)	✓
Verbal Memory-3	✓	✓	✓	✓	✓	✓
Arm Coordination	✓	✓	✓	✓	✓	✓
Draw-A-Child	✓	✓	✓	✓	(✓)	✓
PIPS	-	✓	✓	✓	✓	✓
POCL	✓	✓	-	-	✓	✓

✓ = Acceptable

(✓) = Provisionally acceptable

\- = Not examined

[a]Determined in earlier analyses of spring 1976 data.

The PDC Classroom Observation System. The Observation System was not examined against the same criteria used for the other tests, partly because it is considered to provide measures of group, rather than individual, behavior. By the criteria developed for evaluating the Observation System, necessarily less rigorous than those used for the more conventional instruments, it appears acceptable as a means of assessing the classroom environment. This potential application makes the instrument particularly useful, since it is the classroom that mediates between PDC-induced administrative change (measured in the course of the Implementation Study) and the program's intended effects on children (measured by the child tests). The Observation System provides the only way of examining events at this critical mediating stage. And with the cancellation of the Teacher Survey, this function becomes even more critical.

Other measures. The spring 1977 battery will include all the aforementioned measures plus height and weight (included as pre-measures in this fall's battery) and the PDC Child Rating Scale, which was not administered in the fall because the teachers who complete it cannot be expected to be fully acquainted with the children that early in the year.

Are the PDC and Comparison Groups Really Comparable?

Overall findings. At the site level, Cohort 2 PDC and comparison groups seem to be more similar than were Cohort 1 groups on the critical dimensions of ethnicity, prior preschool experience, and socioeconomic status (represented by mother's education and number of siblings). For half the sites, there were no significant PDC-comparison differences at all on either the background characteristics or the tests. For the other half of the sites, one or two differences each were found, but what's more important is that the PDC and comparison groups in these sites were found to be alike on the remaining 18 or so dimensions examined. The aggregated English-dominant PDC and comparison groups are quite comparable on all the critical background characteristics; the groups differ on only one of thirteen child measures and on none of the six background variables. The aggregated Spanish-dominant groups do not differ on any of the child measures, but do differ in prior preschool experience.

Prospects for analysis at the aggregate level. The composition of the aggregated English-dominant PDC and comparison groups seems quite satisfactory for analytic purposes. In future analyses of test score gain, the one variable on which the groups differed initially can be adjusted without difficulty to make allowances for differences in initial status.

The aggregated Spanish-dominant groups also show considerable similarity, differing significantly on only one of 19 dimensions examined. But that one is prior preschool experience, a factor that might well be expected to make a difference in children's adjustment to school life, thus possibly obscuring whatever effects PDC may produce. (Only 11% of the Spanish-dominant PDC children attended preschool before the present Head Start year, versus 39% of the Spanish-dominant comparison children.) Still, despite this speculation, it may turn out that the difference in preschool experience will actually be of little consequence (both groups, after all, will have had at least one year of Head Start). And even if prior preschool experience is found in future years to affect the performance of these children, certain statistical strategies can be applied when analyzing PDC's effects to reduce the bias introduced by the differing preschool experience of the PDC and comparison groups.

Thus considering the two major possibilities--(a) that the difference in prior preschool experience may have no future biasing effect and (b) that even if a biasing effect is discovered it can probably be minimized statistically--it is recommended that the evaluation proceed for the Spanish-dominant sample. This aspect of the evaluation, it should be noted, is of special interest, since bilingual/bicultural education is one of PDC's special emphases.

Prospects for analysis at the site level. The value of performing analyses of child impact at the site level is questionable for two reasons. First, there are indications of some initial group differences at about half of the sites. Second, there may be less to learn from twelve separate site-level analyses than there is from analyses that examine groups aggregated across sites; this is so because site-level analyses command less statistical power and because it is difficult to draw generalizations from findings that may fluctuate unsystematically from site to site.

Will Large Enough Samples of Children Remain in PDC and Comparison Schools at each Site to Permit a Longitudinal Study of Program Effects?

It is obvious from attrition projections that if the evaluation depended upon site-level analyses of PDC's effects on children, the sample sizes available at most sites would be inadequate by the time Cohort 2 reaches first grade; it is possible that as soon as next year the number of Cohort 2 children remaining at PDC and comparison schools in some sites will be too small to permit statistically adequate within-site analyses. By aggregating PDC and comparison groups across sites, however, a sufficient sample can be constituted to allow analyses to continue through 1980-81, when Cohort 2 will be in grade 3. This is certainly true for the English-dominant sample, at least. It is less certainly true for the Spanish-dominant sample. However, even for the latter, analyses could proceed for a few years--long enough to allow preliminary conclusions to be drawn about the effects of PDC.

The issue is largely one of cost: are the projected final samples of English-dominant PDC and comparison children--numbering about 200 and 170, respectively--large enough to justify the expense of the Impact Study? The question cannot be answered fully within this report--a wider forum is required.

Prospects for Testing OCD's Hypotheses Regarding PDC's Impact on Children

OCD's original specifications for the PDC evaluation included three hypotheses concerning the program's expected impact on children; since the time when that document was written a fourth hypothesis has been developed. Testing of these hypotheses, which are recapitulated below, will be the major mission of the Impact Study if it continues longitudinally.

1. *Head Start-Entry PDC Group vs. Head Start-Entry Comparison Group:* Children who enter a PDC program at the Head Start level will show significant gains (on tests in the social competence battery) over comparable children who enter a non-PDC program at the Head Start level.

2. *Head Start-Entry PDC Group vs. Kindergarten-Entry PDC Group:* Children who enter a PDC program at the Head Start level will show significant gains over comparable age-mates who enter the PDC program in kindergarten, not having attended Head Start.

3. *Kindergarten-Entry PDC Group vs. Head Start-Entry Comparison Group:* Children who enter PDC at the kindergarten level will show significant gains over comparable age-mates who enter a non-PDC Head Start and progress through a non-PDC elementary school.

4. *Kindergarten-Entry PDC Group vs. Kindergarten-Entry Comparison Group:* Children who enter a PDC program at the kindergarten level will show significant gains over comparable children who enter a comparison program at the kindergarten level.

Hypothesis 1 is PDC's cornerstone, and up to the present, the efforts of the Impact Study have been directed solely toward determining the feasibility of testing Hypothesis 1. The conclusions drawn in the preceding section affirm that the hypothesis can indeed by tested with a reasonable expectation of obtaining decisive results. Thus it is appropriate now to consider the feasibility of testing Hypotheses 2, 3, and 4.

Testing of Hypotheses 2 and 3 will require identification of a sample of children who (a) resemble Cohort 2 PDC children on educationally relevant dimensions, (b) have had no Head Start experience, and (c) enter PDC at the kindergarten level. Testing of Hypothesis 4 will require identification of another group of children who satisfy conditions (a) and (b) and who enter a <u>comparison</u> school at the kindergarten level.

The feasibility of testing these hypotheses depends upon the chances of being able to locate children who are like Cohort 2 children on the dimensions examined in this report (e.g., ethnicity, family background), who would have been eligible, or nearly eligible, to attend Head Start, but did not, and who attend the same schools as Cohort 2 children. These conditions present several problems, enumerated below.

(1) It is far from certain that such children exist in numbers sufficient for statistical investigation.

(2) If there are enough such children, it will be difficult to identify them because the information required to establish socioeconomic status, which is one of the major matching criteria, is often inaccessible, or is accessible only at great cost in time and effort.

(3) If the children do exist and are identified, the problem remains of establishing a baseline for their performance on the measures in the PDC battery.

The third point calls for amplification. Hypotheses 2 and 3 require that children who enter a PDC program in kindergarten be compared with children of the same age who entered a PDC (or comparison) program at the Head Start level. The mechanics of the analytic procedure in this case will actually involve comparing the progress of one group of children with the progress of the other. And to measure progress, it is necessary to determine each child's baseline at the onset of the program. For Cohort 2 children, the program began at age 4 with the Head Start year, and each child's performance baseline was determined by administration of the fall 1976 battery. But since the kindergarten-entry children cannot be identified until they enter school at age 5, it will not be possible to establish an age 4 baseline for them directly. Of course, it is possible to use age 5 as a baseline for both groups, but the risk in this is that PDC or Head Start will already have raised the performance level of the Cohort 2 children, and treating age 5 as the starting point will render the program effect invisible.

This discussion is not intended to discourage a decision to pursue the testing of Hypotheses 2 through 4; its purpose is to delineate the issues that must be considered before such a decision is made. Final recommendations on this and other matters will be presented to OCD immediately after the April meeting of the PDC Evaluation Advisory Panel.

APPENDIX A

Descriptions of the Measures in the Fall Battery

Descriptions of the Measures in the Fall Battery

	Order of Administration[1]
• **Social-Emotional Measures**	
PDC Classroom Observation System	1
Preschool Interpersonal Problem-Solving Test (PIPS)	7
Pupil Observation Checklist (POCL)	9
• **Psychomotor Measures**	
Arm Coordination [McCarthy Scales of Children's Ability (MSCA)]	8
• **Cognitive and Language Measures**	
Bilingual Syntax Measure (BSM)	2
Block Design (WPPSI)	4
Draw-A-Child (MSCA)	5
Verbal Memory (MSCA)	3
Verbal Fluency (MSCA)	6
• **Other Measures**	
Adult Language Check	
Attrition, Handicap and Attendance Information Sheet	

Each of these measures is described briefly below. For a more extensive review, see Interim Report II, Part B: Recommendations for Measuring Program Impact (1975).

---

[1]As noted in the text, the battery was administered in two or sometimes three sessions.

PDC Classroom Observation System (High/Scope Foundation, unpublished). The PDC observation system was developed to provide information about children's classroom behavior along dimensions pertinent to the social-emotional goals of Project Developmental Continuity. The system focuses on aspects of an individual child's behavior, verbal or nonverbal, that reflect the child's attitude toward himself, and on the child's social competence as demonstrated in his interaction with peers and adults.

Using a time sampling method, trained observers observe each child for five minutes at four different times during the day and code their behavior into four general categories: "noninvolved," "involved," "interacts with peer," and "interacts with adult." A fifth category, "activity level," is included to provide information concerning the context in which these behaviors were observed. Each of these categories includes subcategories that are designed to identify the frequency and nature of specific behaviors within the general category.

Preschool Interpersonal Problem-Solving Test (Shure and Spivack, 1974)[1]. The PIPS attempts to assess the child's ability to name alternative solutions to a life-related problem--that of obtaining a toy from another child. Paper cut-outs of boys, girls and toys are used in presenting the problem. Among inner city four-year-olds attending the Philadelphia Get Set day care program, those judged as better-adjusted by their teachers were able to conceptualize a greater number and a wider range of alternative solutions to real-life problems than were their more poorly adjusted classmates.

Pupil Observation Checklist (High/Scope Foundation, unpublished). This is a rating scale consisting of twelve 7-point bipolar adjectives derived from a similar scale used in the Home Start evaluation.[2] The tester rates each child using this instrument after he or she has administered all the other measures in the battery to the child. See Appendix H for details on the factor structure of this instrument.

---

[1]Shure, M. B. & Spivack, G. The PIPS Test Manual. Philadelphia: Hahneman Medical College, 1974.

[2]Love, J.,et al. National Home Start Evaluation Interim Report VII. Ypsilanti, MI: High/Scope Foundation, March 1976.

McCarthy Scales of Children's Abilities (McCarthy, 1972)[1]. These subtests consist of a series of tasks tapping problem-solving, psychomotor, and conceptual abilities, and are similar to the Wechsler scales, but with emphasis on age-related maturational indicators.

- Verbal Memory. The child is asked to repeat sequences of words (Verbal Memory-1) and to repeat or retell as much as possible of a one paragraph story (Verbal Memory-3).

- Verbal Fluency. The child is asked to name as many members of specific categories (e.g., animals) as he/she can.

- Arm Coordination. Child bounces a rubber ball, catches a beanbag, and throws a beanbag through a hole in a target.

- Draw-A-Child. Child draws a picture of a child of the same sex.

Wechsler Preschool and Primary Scale of Intelligence, Block Design subtest (Wechsler, 1967)[2]. The task requires reproducing (constructing) designs with flat colored blocks, either from the examiner's model or from a picture on a card. The measure taps problem-solving abilities, flexibility of response style, visual-motor organization, and execution.

Bilingual Syntax Measure (Burt, Dulay and Hernandez-Ch., 1975)[3]. This test is designed to measure children's oral proficiency in English and/or Spanish standard grammatical structures. Simple questions are used with cartoon-type colored pictures to provide a conversational setting for eliciting natural speech. An analysis of the child's response yields a numerical indicator and a qualitative description of the child's structural language proficiency in standard English or standard Spanish. Responses are written down verbatim. The English version is administered only to English-dominant or bilingual children; the Spanish version only to Spanish-dominant or bilingual children.

---

[1]McCarthy, D. McCarthy Scales of Children's Abilities: Manual. New York: Psychological Corporation, 1972.

[2]Wechsler, D. Wechsler Preschool and Primary Scale of Intelligence: Manual. New York: Psychological Corporation, 1967.

[3]Burt, M., Dulay, H. & Hernandez-Ch., E. Bilingual Syntax Measure. New York: Harcourt, Brace, Jovanovich, 1975.

Adult Language Check. This measure is used in the bilingual/bicultural demonstration sites to obtain an indication of the languages the adults in the classroom use during their interactions with children. The interviewer sits in the classroom for a two-hour period and records the language used by the teacher, and aides approximately every five minutes. The Adult Language Check was used only in classrooms where languages other than English were spoken by teachers, aides, or other adults.

Demographic Information Sheet. Additional information about each child in the sample, such as previous preschool experience, handicap status, dominant language, etc. is obtained from Head Start records.

Height and Weight. All children are weighed and measured during the same two-week period in the fall.

Controlling for Order Effect in Administering the BSM to Bilingual Children

Children who show facility in both Spanish and English receive both versions of the BSM. The order in which the two versions are administered is controlled so that during any single testing period half the children receive the Spanish version first and half receive the English version first. Further, the order is reversed with each successive testing so that, for example, a child who received the Spanish version first in the fall would receive the English version first in the spring.

APPENDIX B

Forms for Weekly Tester Monitoring

The forms reproduced here were used weekly by testers for mutual monitoring. The completed forms were returned regularly to High/Scope for continuing analysis. In this appendix, the categories beside which an X appears are those in which testers, as a group made more errors than expected or than was judged tolerable.

Table B-1
ARM COORDINATION

Monitoring Form

Interviewer _____ Date _____

Child's Name _____

INSTRUCTIONS: This form will provide High/Scope Foundation with information on how similar the interview administrations are within each site and across sites. The interviews must be administered in a standard or uniform way to insure comparability of the data. When you monitor another interviewer you should be recording the child's responses in your interview booklet and be watching for and noting whether any of the following errors occur during each of the interviews. You will fill out one of these monitoring forms for <u>each</u> interview you monitor.

Test Administration Errors	Check Each Time Error Occurs
1. Fails to have CORRECT INTERVIEWING MATERIALS; e.g., didn't have ball, beanbag, tape, etc.	_____
2. INCORRECT PLACEMENT of interview materials; e.g., didn't have target 6' from child, didn't kneel or bend when throwing beanbag to child, etc.	_____
3. INCORRECT WORDING of interview questions; e.g., doesn't follow the words in the interview booklet.	_____
4. SKIPPED AN ITEM.	_____
5. SKIPPED A SECOND TRIAL, or gave a second trial when it should not have been given.	_____
6. STOPPED INTERVIEW INCORRECTLY; e.g., gave Part B in Beanbag Catch Game when child failed to catch in Part A.	_____
7. REPEATS; repeated the interview question more than one time.	_____
8. ENCOURAGEMENTS; gave more than one encouragement per initial question and repeat; didn't give an encouragement when needed.	_____
9. SCORING; scored child's response incorrectly.	_____
10. INCORRECT TIMING; failed to mark time stopped on test booklet.	_____
11. OTHER: (specify) _____	
_____	
_____	

Rapport with child (circle one): Poor Adequate Good

Name of Monitor _____

Table B-2

BILINGUAL SYNTAX MEASURE

Monitoring Form

Interviewer _____ **Date** _____

Child's Name _____

INSTRUCTIONS: This form will provide High/Scope Foundation with information on how similar the interview administrations are within each site and across sites. The interviews must be administered in a standard or uniform way to insure comparability of the data. When you monitor another interviewer you should be recording the child's responses in your interview booklet and be watching for and noting whether any of the following errors occur during each of the interviews. You will fill out one of these monitoring forms for <u>each</u> interview you monitor.

Test Administration Errors	Check Each Time Error Occurs
1. Fails to have CORRECT INTERVIEWING MATERIALS; e.g., is missing the warm-up picture.	_____
2. INCORRECT PLACEMENT of interview materials; e.g., doesn't place warm-up picture directly in front of child, doesn't place picture booklet directly in front of child; didn't put warm-up picture out of child's sight when using booklet, etc.	_____
3. INCORRECT WORDING of interview questions; e.g., doesn't follow the words in the interview booklet, adds too many additional comments or questions.	_____
4. SKIPPED AN ITEM.	_____
5. STOPPED INTERVIEW INCORRECTLY; e.g., didn't stop after item 5 when child responded to only two of the first five items; didn't stop after four DK-R-NR.	_____
6. INCORRECT TIMING; e.g., didn't mark time started and time stopped on cover of interview booklet.	_____
7. REPEATS; repeated the interview question more than one time or didn't repeat the question when it should have been repeated; repeated the child's response verbally.	_____
8. ENCOURAGEMENTS; gave more than one encouragement after the initial question; gave more than one encouragement after the repeat or didn't give an encouragement when it should have been given.	_____
9. SCORING; not writing child's response <u>exactly</u> as said; not writing legibly.	_____
10. DEFINES WORDS; defining words for child during the non-preliminary questions.	_____
11. OTHER: (specify) _____ _____ _____	

Rapport with child (circle one): Poor Adequate Good

Name of Monitor _____

Table B-3

DRAW-A-CHILD

Monitoring Form

Interviewer _____ Date _____

Child's Name _____

INSTRUCTIONS: This form will provide High/Scope Foundation with information on how similar the interview administrations are within each site and across sites. The interviews must be administered in a standard or uniform way to insure comparability of the data. When you monitor another interviewer you should be recording the child's responses in your interview booklet and be watching for and noting whether any of the following errors occur during each of the interviews. You will fill out one of these monitoring forms for each interview you monitor.

Test Administration Errors	Check Each Time Error Occurs
1. INCORRECT PLACEMENT of interview materials; e.g., didn't place blank page width-wise in front of child.	_____
2. INCORRECT WORDING of interview questions; e.g., doesn't follow the words in the interview booklet.	_____
3. INCORRECT TIMING; e.g., didn't mark time started.	_____
4. REPEATS; repeats the interview question.	_____
5. ENCOURAGEMENTS; failed to give one encouragement specified in test booklet when needed.	_____
6. OTHER: (specify) _____	
_____	
_____	

Rapport with child (circle one): Poor Adequate Good

Name of Monitor _____

Table B-4

PIPS

Monitoring Form

Interviewer _____ Date _____

Child's Name _____

INSTRUCTIONS: This form will provide High/Scope Foundation with information on how similar the interview administrations are within each site and across sites. The interviews must be administered in a standard or uniform way to insure comparability of the data. When you monitor another interviewer you should be recording the child's responses in your interview booklet and be watching for and noting whether any of the following errors occur during each of the interviews. You will fill out one of these monitoring forms for each interview you monitor.

Test Administration Errors	Check Each Time Error Occurs
1. Fails to have CORRECT INTERVIEWING MATERIALS; e.g., missing one of the PIPS cutouts, etc.	_____
2. INCORRECT PLACEMENT of interview materials; e.g., putting toy on wrong cut-out, placing cut-outs on table rather than on some kind of stand.	_____
3. INCORRECT WORDING of interview questions; e.g., doesn't follow the words in the interview booklet.	_____
4. SKIPPED AN ITEM.	_____
5. STOPPED INTERVIEW INCORRECTLY; e.g., didn't stop interview after two consecutive stories in which child gave repetition of answers, no solution answers, or DK-R-NR.	_____
6. PROBING, too many or too few; e.g., didn't probe when response required it or probed when child's answer was acceptable.	_____
7. SCORING; recorded child's response incorrectly; failed to put child's response in correct response box.	X
8. OTHER: (specify) _____	
_____	
_____	

Rapport with child (circle one): Poor Adequate Good

Name of Monitor _____

Table B-5

VERBAL FLUENCY
Monitoring Form

Interviewer _____ **Date** _____

Child's Name _____

INSTRUCTIONS: This form will provide High/Scope Foundation with information on how similar the interview administrations are within each site and across sites. The interviews must be administered in a standard or uniform way to insure comparability of the data. When you monitor another interviewer you should be recording the child's responses in your interview booklet and be watching for and noting whether any of the following errors occur during each of the interviews. You will fill out one of these monitoring forms for <u>each</u> interview you monitor.

Test Administration Errors	Check Each Time Error Occurs
1. INCORRECT WORDING of interview questions; e.g., doesn't follow the words in the interview booklet.	_____
2. SKIPPED AN ITEM.	_____
3. STOPPED INTERVIEW INCORRECTLY; e.g., didn't give entire interview.	_____
4. INCORRECT TIMING; e.g., allowed the child more than or less than 20 seconds to name all the toys he could think of (tester and monitor should be within 5 seconds of each other on the timing).	_____
5. REPEATS; repeated the interview question.	_____
6. SCORING; didn't record child's response <u>exactly</u> as said, didn't write legibly.	_____
7. ENCOURAGEMENTS; failed to say appropriate encouragement after 5 seconds when it should have been said, or encouraged too many times.	_____
8. OTHER: (specify) _____	
_____	
_____	

Rapport with child (circle one): Poor Adequate Good

Name of Monitor _____

Table B-6

VERBAL MEMORY

Monitoring Form

Interviewer _____ Date _____

Child's Name _____

INSTRUCTIONS: This form will provide High/Scope Foundation with information on how similar the interview administrations are within each site and across sites. The interviews must be administered in a standard or uniform way to insure comparability of the data. When you monitor another interviewer you should be recording the child's responses in your interview booklet and be watching for and noting whether any of the following errors occur during each of the interviews. You will fill out one of these monitoring forms for each interview you monitor.

Test Administration Errors	Check Each Time Error Occurs
1. INCORRECT WORDING of interview questions; e.g., doesn't follow the words in the interview booklet.	_____
2. SKIPPED AN ITEM.	_____
3. STOPPED INTERVIEW INCORRECTLY; e.g., failed to stop Part I after child gave no correct answers to items 2 and 3, or items 4 and 5; gave Part II when shouldn't have.	_____
4. INCORRECT TIMING; didn't mark time started on test booklet.	_____
5. REPEATS; repeated the interview question.	_____
6. SPEED; read the words too quickly for the child or allowed too much time between the words.	_____
7. ENCOURAGEMENTS; encouraged the child more than once or didn't encourage the child at all when he didn't respond or didn't use specified encouragement.	X
8. SCORING; failed to record child's response correctly or wrote child's response illegibly.	_____
9. OTHER: (specify) _____	
_____	
_____	

Rapport with child (circle one): Poor Adequate Good

Name of Monitor _____

Table B-7

WPPSI BLOCK DESIGN

Monitoring Form

Interviewer _____ Date _____

Child's Name _____

INSTRUCTIONS: This form will provide High/Scope Foundation with information on how similar the interview administrations are within each site and across sites. The interviews must be administered in a standard or uniform way to insure comparability of the data. When you monitor another interviewer you should be recording the child's responses in your interview booklet and be watching for and noting whether any of the following errors occur during each of the interviews. You will fill out one of these monitoring forms for each interview you monitor.

Test Administration Errors	Check Each Time Error Occurs
1. Fails to have CORRECT INTERVIEWING MATERIALS; e.g., doesn't have all 14 blocks, doesn't have picture booklet.	_____
2. INCORRECT PLACEMENT of interview materials; e.g., makes incorrect WPPSI design, uses wrong blocks in making design, giving wrong blocks to child.	_____
3. INCORRECT WORDING of interview questions; e.g., doesn't follow the words in the interview booklet.	_____
4. SKIPPED AN ITEM.	_____
5. SKIPPED A SECOND TRIAL, or gave a second trial when it should not have been given.	_____
6. STOPPED INTERVIEW INCORRECTLY; e.g., didn't stop after two consecutive failures, or didn't give item 4 after child failed items 2 and 3.	_____
7. INCORRECT TIMING; e.g., allowed the child more or less time to make the design than the instructions indicated (tester and monitor should be within 5 seconds of each other on the timing), didn't mark time stopped.	_____
8. REPEATS; gave a demonstration when it should not have been given or failed to give a demonstration; repeated the interview question.	_____
9. ENCOURAGEMENTS; gave more than one encouragement or none at all with the initial question and more than one encouragement or none at all on the second trial.	_____
10. SCORING; scored child's response incorrectly.	_____
11. ROTATIONS and GAPS; failed to correct child's rotations or ask child, "Is that right?" when he left more than 1/4 inch between his blocks.	_____
12. OTHER: (specify) _____ _____	

Rapport with child (circle one): Poor Adequate Good

Name of Monitor:

APPENDIX C

Commentary on Scoring the McCarthy Arm Coordination Scale

Commentary on Scoring the McCarthy Arm Coordination Scale

Following the fall 1975 data collection, a number of alternative procedures were explored for scoring items and scale scores in the child test battery. Particular attention was paid to the subtests of the McCarthy Scales of Children's Abilities (Draw-A-Child, Verbal Fluency, Verbal Memory, and Arm Coordination). These efforts were reported in Interim Report III, Part A (March 1, 1976). The conclusion at that time was that since there were no apparent differences between the results of McCarthy scoring procedures and alternative procedures, there would be no advantage in adhering to the McCarthy scoring conventions.

However, re-examination of the psychometric data for the three administrations of the battery indicates a need for revising this position with respect to the Arm Coordination scale. When scored using McCarthy criteria, internal consistency reliability (alpha) coefficients have been consistently lower than those for other tests in the battery (for the English-dominant sample, .54, .62, and .65 for fall 1975, spring 1976, and fall 1976). Since this scale is the only "pure" psychomotor measure in the battery, the incentive to retain it is greater than it would be if it were redundant with other constructs tapped by the battery. Thus alternative scoring procedures were explored once again.

It was found that internal consistency coefficients for the fall 1976 Arm Coordination data were somewhat higher when each of the six items in the scale as weighted equally than when they were weighted unequally, as they are in the McCarthy procedures. (Logic also argues for equal weighting since no argument is known to exist for unequal weighting.) Although the internal consistency coefficient for the English-dominant sample still just manages to reach the nominal criterion value of .65, it is judged to be sufficient to warrant retention of this instrument, especially when consideration is given to the relatively high test-retest correlation, .72, found last year.

APPENDIX D

Flowcharts for the Analysis Procedure

Figure D-1

Flow Chart for Step 1: Is the Measure Reliable for This Sample?

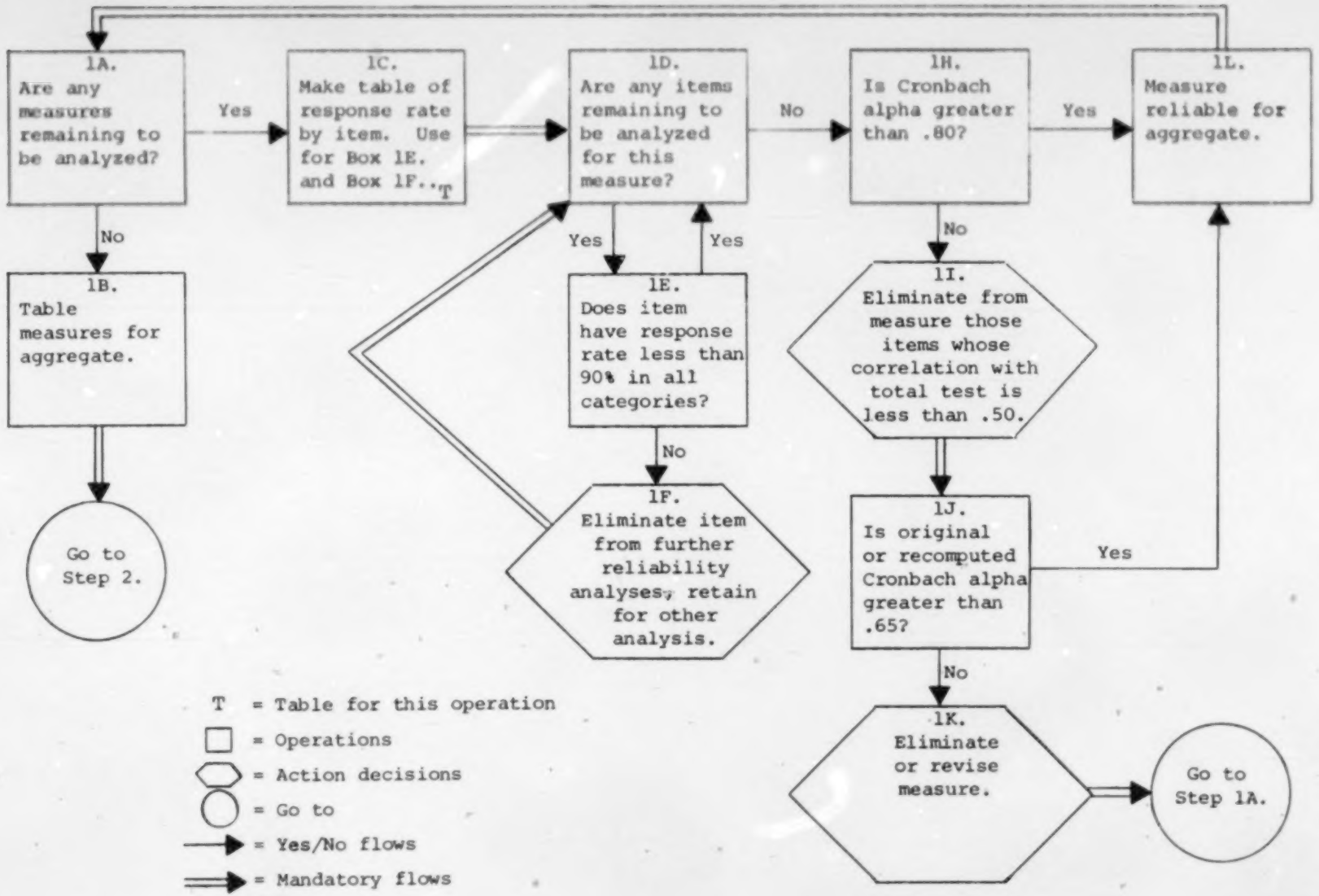

1A. Are any measures remaining to be analyzed?
Yes
1C. Make table of response rate by item. Use for Box 1E. and Box 1F.. T
1D. Are any items remaining to be analyzed for this measure?
No
1H. Is Cronbach alpha greater than .80?
Yes
1L. Measure reliable for aggregate.
No
1B. Table measures for aggregate.
Go to Step 2.
Yes
Yes
1E. Does item have response rate less than 90% in all categories?
No
1F. Eliminate item from further reliability analyses, retain for other analysis.
No
1I. Eliminate from measure those items whose correlation with total test is less than .50.
1J. Is original or recomputed Cronbach alpha greater than .65?
Yes
No
1K. Eliminate or revise measure.
Go to Step 1A.
T = Table for this operation
= Operations
= Action decisions
= Go to
= Yes/No flows
= Mandatory flows

Figure D-2

Flow Chart for Step 2: Is the Measures Valid for This Sample?

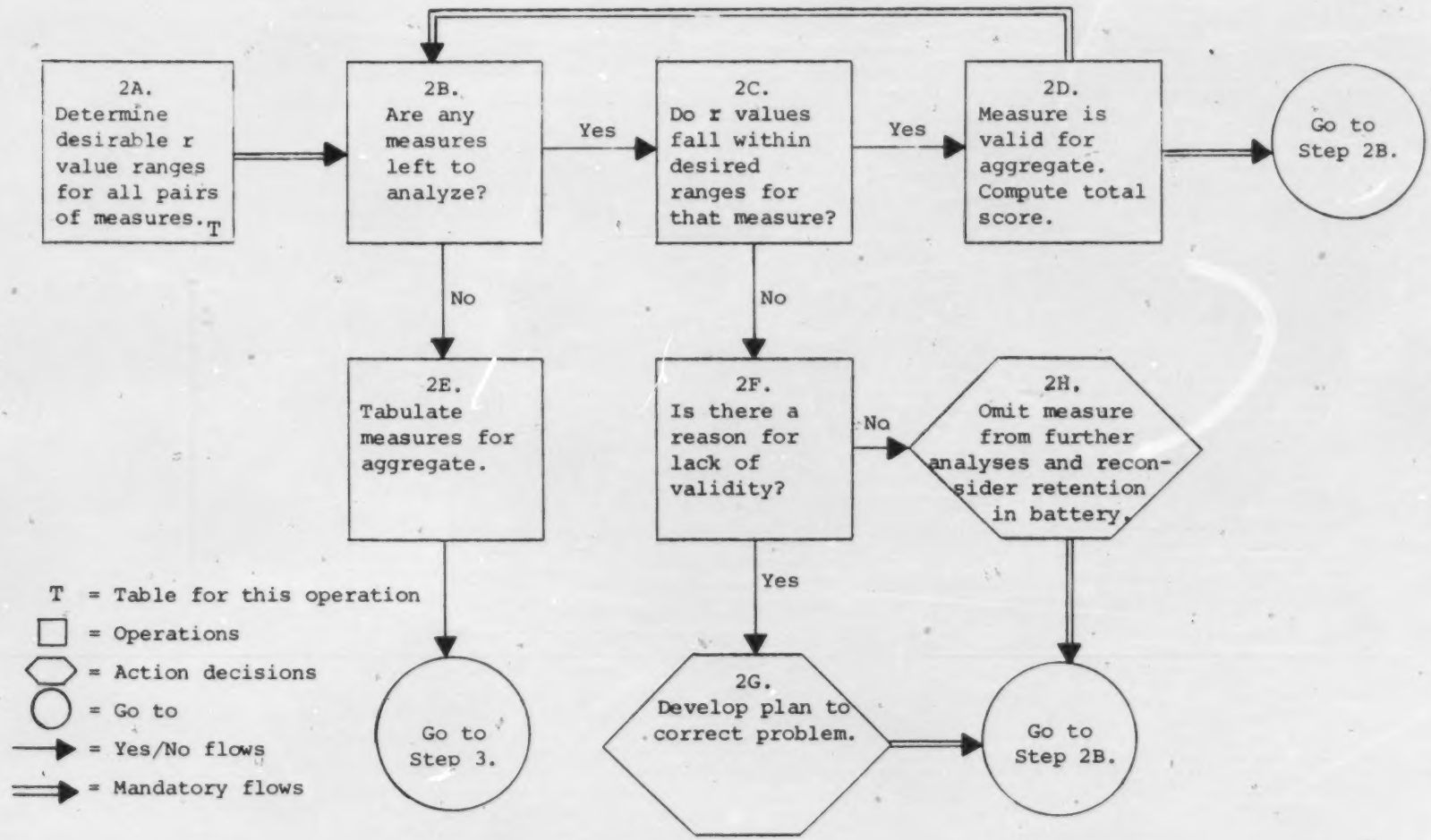

2A. Determine desirable r value ranges for all pairs of measures. T
2B. Are any measures left to analyze?
Yes
2C. Do r values fall within desired ranges for that measure?
Yes
2D. Measure is valid for aggregate. Compute total score.
Go to Step 2B.
No
No
2E. Tabulate measures for aggregate.
2F. Is there a reason for lack of validity?
No
2H. Omit measure from further analyses and reconsider retention in battery.
Yes
2G. Develop plan to correct problem.
Go to Step 2B.
Go to Step 3.
T = Table for this operation
= Operations
= Action decisions
= Go to
= Yes/No flows
= Mandatory flows

Figure D-3

Flow Chart for Step 3: Are Reliability and Validity Constant Across Time and Samples?

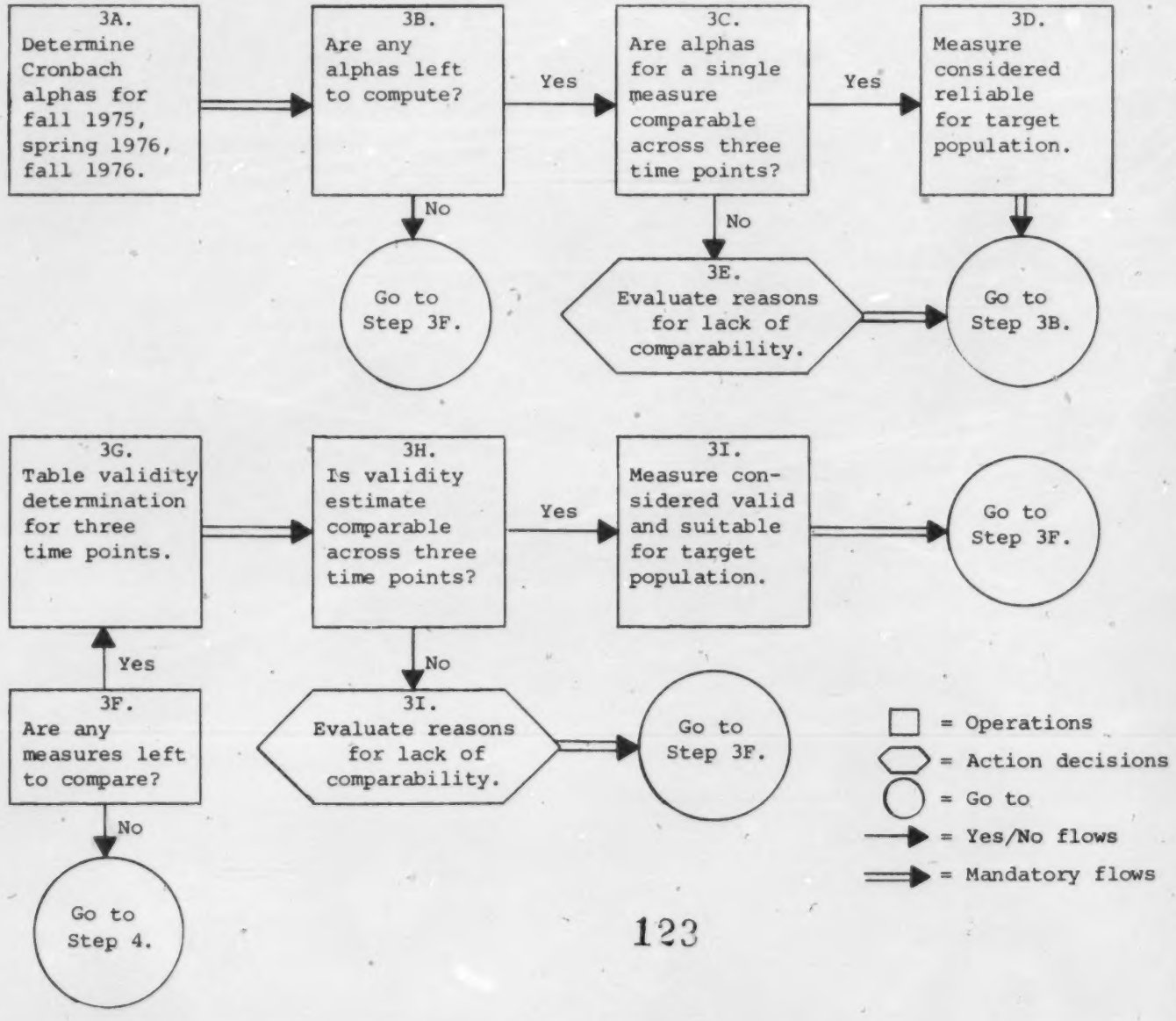

3A. Determine Cronbach alphas for fall 1975, spring 1976, fall 1976.
3B. Are any alphas left to compute?
Yes
No
Go to Step 3F.
3C. Are alphas for a single measure comparable across three time points?
Yes
No
3E. Evaluate reasons for lack of comparability.
Go to Step 3B.
3D. Measure considered reliable for target population.
3G. Table validity determination for three time points.
3H. Is validity estimate comparable across three time points?
Yes
No
3I. Measure considered valid and suitable for target population.
Go to Step 3F.
3I. Evaluate reasons for lack of comparability.
Go to Step 3F.
3F. Are any measures left to compare?
Yes
No
Go to Step 4.
= Operations
= Action decisions
= Go to
= Yes/No flows
= Mandatory flows

Figure D-4

Flow Chart for Step 4: Does the Factor Structure Support Expectations?

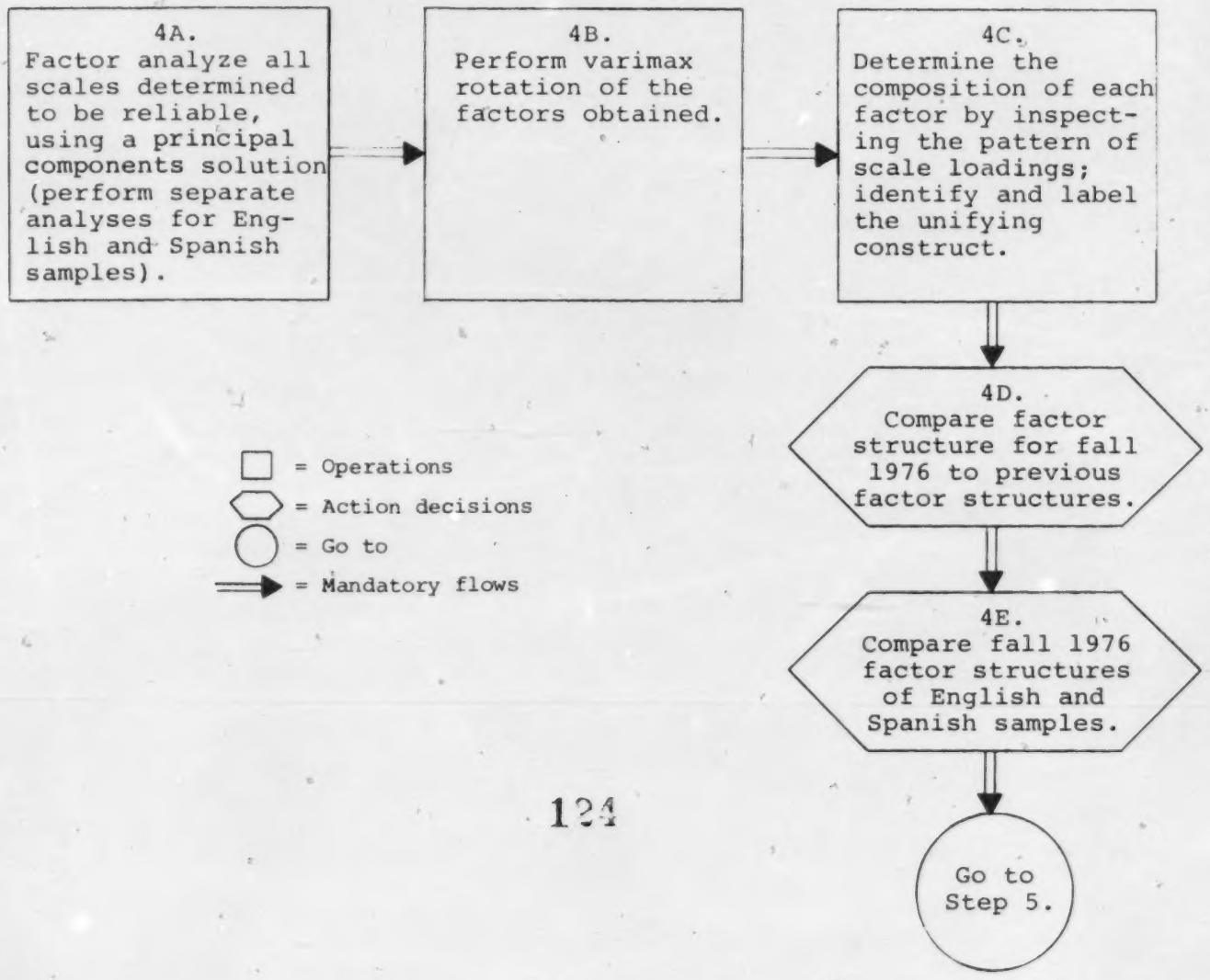

4A.
Factor analyze all scales determined to be reliable, using a principal components solution (perform separate analyses for English and Spanish samples).
4B.
Perform varimax rotation of the factors obtained.
4C.
Determine the composition of each factor by inspecting the pattern of scale loadings; identify and label the unifying construct.
4D.
Compare factor structure for fall 1976 to previous factor structures.
4E.
Compare fall 1976 factor structures of English and Spanish samples.
Go to Step 5.
= Operations
= Action decisions
= Go to
= Mandatory flows

Figure D-5

Flow Chart for Step 5:
Are the PDC and Comparison Groups Comparable?

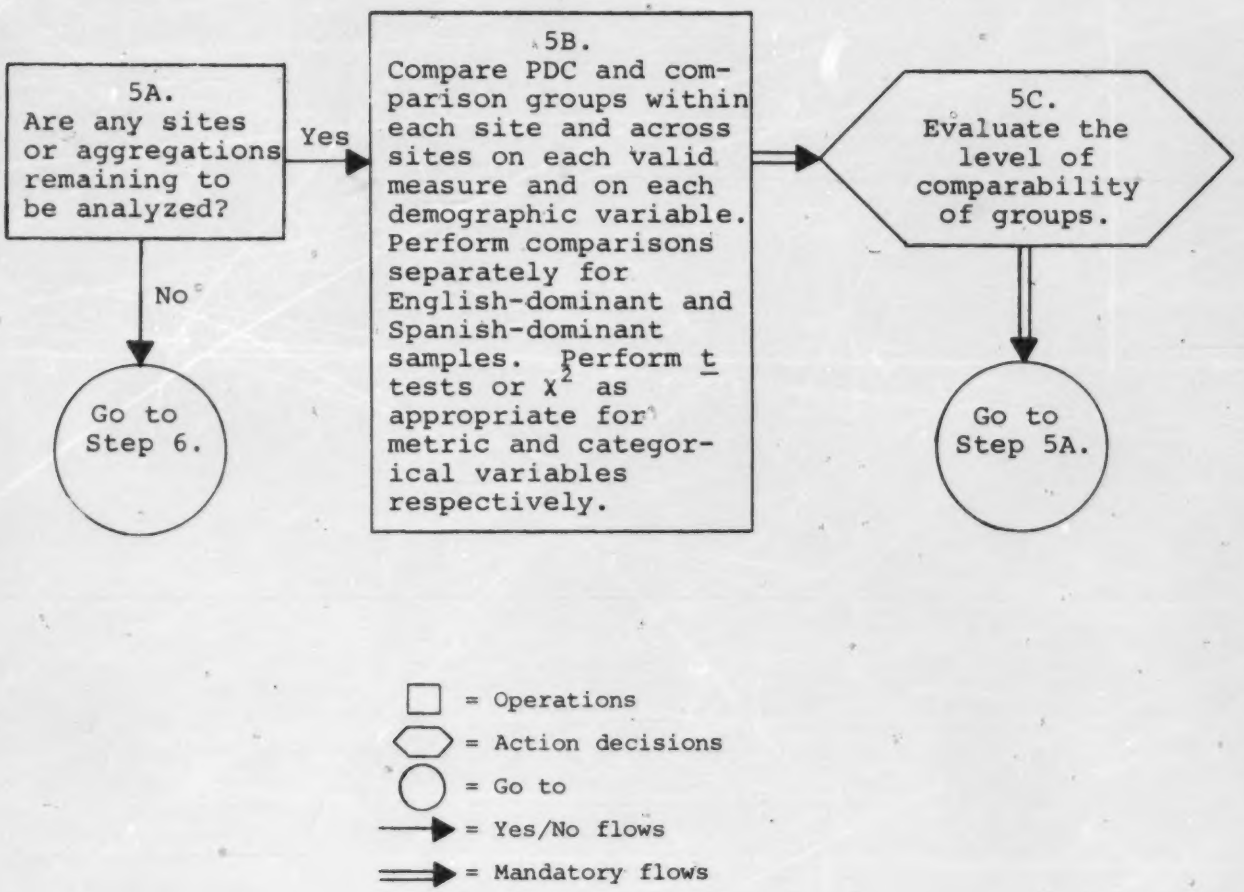
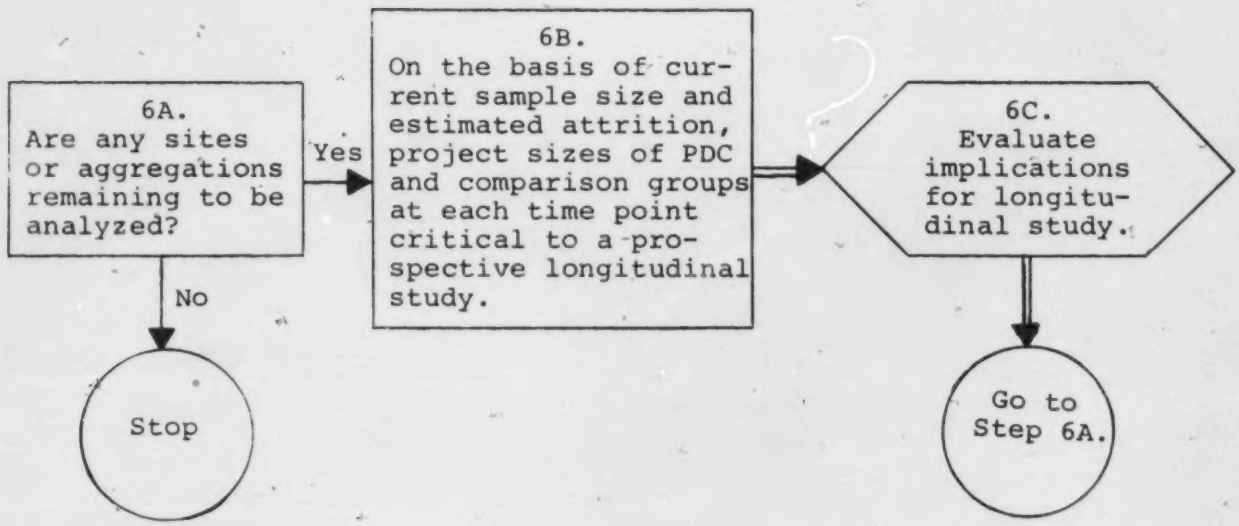
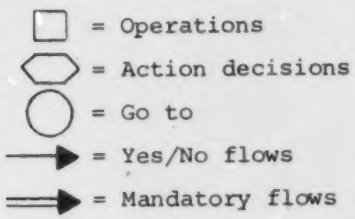

5A.
Are any sites or aggregations remaining to be analyzed?
Yes
No
Go to Step 6.
5B.
Compare PDC and comparison groups within each site and across sites on each valid measure and on each demographic variable. Perform comparisons separately for English-dominant and Spanish-dominant samples. Perform t tests or χ^2 as appropriate for metric and categorical variables respectively.
5C.
Evaluate the level of comparability of groups.
Go to Step 5A.

Figure D-6

Flow Chart for Step 6: Are Sample Sizes and Retention Rates Adequate?


6A.
Are any sites or aggregations remaining to be analyzed?
Yes
No
Stop
6B.
On the basis of current sample size and estimated attrition, project sizes of PDC and comparison groups at each time point critical to a prospective longitudinal study.
6C.
Evaluate implications for longitudinal study.
Go to Step 6A.



APPENDIX E

Magnitude of Differences for Variables on Which Groups Were Found Unequal, by Site

Table E-1

California-English

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N≈37)	Comparison (N≈24)
Background Characteristics		
No differences found.		
Test Performance		
No differences found.		

Table E-2

California-Spanish

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N=7)	Comparison (N=15)
Background Characteristics		
No differences found.		
Test Performance		
No differences found.		

Table E-3

Colorado

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N=51)	Comparison (N=30)
Background Characteristics		
No differences found.		
Test Performance		
No differences found.		

Table E-4

Connecticut

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N=37)	Comparison (N=55)
Background Characteristics		
No differences found.		
Test Performance		
No differences found.		

Table E-5

Georgia

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N=43)
No comparison group.	

Table E-6

Florida

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N=45)	Comparison (N=49)
Background Characteristics		
No differences found.		
Test Performance		
POCL-2	-.55z	.24z

Note

Group means are given as z scores. A z score expresses the difference between one group's mean and the mean for all cases in terms of standard deviation units (e.g., .50z represents a position 1/2 standard deviation above the overall mean).

Table E-7

Iowa

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N=49)	Comparison (N=51)
Background Characteristics		
No differences found.		
Test Performance		
Verbal Fluency	-.34z	.34z
Verbal Memory-1	-.33z	.32z

Note

Group means are given as z scores. A z score expresses the difference between one group's mean and the mean for all cases in terms of standard deviation units (e.g., .50z represents a position 1/2 standard deviation above the overall mean).

Table E-8

Maryland

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N=32)	Comparison (N=45)
Background Characteristics		
No differences found.		
Test Performance		
Verbal Memory-3	.40z	-.28z

Note

Group means are given as z scores. A z score expresses the difference between one group's mean and the mean for all cases in terms of standard deviation units (e.g., .50z represents a position 1/2 standard deviation above the overall mean).

Table E-9

Michigan

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N=58)	Comparison (N=58)
Background Characteristics		
No differences found.		
Test Performance		
Verbal Memory-1	-.23z	.24z

Note

Group means are given as z scores. A z score expresses the difference between one group's mean and the mean for all cases in terms of standard deviation units (e.g., .50z represents a position 1/2 standard deviation above the overall mean).

Table E-10

Texas-English

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N=26)	Comparison (N=19)
Background Characteristics		
No differences found.		
Test Performance		
No differences found.		

Table E-11

Texas-Spanish

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N≈38)	Comparison (N≈34)
Background Characteristics		
Age	-.23z	.27z
Prior Preschool		
Yes	3%	33%
No	97%	67%
Test Performance		
No differences found.		

Note

Group means are given as z scores. A z score expresses the difference between one group's mean and the mean for all cases in terms of standard deviation units (e.g., .50z represents a position 1/2 standard deviation above the overall mean).

Table E-12

Utah

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N≈61)	Comparison (N≈55)
Background Characteristics		
No differences found.		
Test Performance		
Height	.23z	-.25z

Note

Group means are given as z scores. A z score expresses the difference between one group's mean and the mean for all cases in terms of standard deviation units (e.g., .50z represents a position 1/2 standard deviation above the overall mean).

Table E-13

Washington

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N≈49)	Comparison (N≈66)
Background Characteristics		
No differences found.		
Test Performance		
No differences found.		

Table E-14

West Virginia

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC (N=42)	Comparison (N=29)
Background Characteristics		
No differences found.		
Test Performance		
No differences found.		

Table E-15

English-Dominant Aggregate

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC	Comparison
Background Characteristics		
No differences found.		
Test Performance		
Verbal Memory-1	-.09z	.10z

Note

Group means are given as z scores. A z score expresses the difference between one group's mean and the mean for all cases in terms of standard deviation units (e.g., .50z represents a position 1/2 standard deviation above the overall mean).

Table E-16

Spanish-Dominant Aggregate

Magnitude of Differences Between PDC and Comparison Analytic Samples for Variables on Which Samples Were Found Unequal ($p<.01$)

	PDC	Comparison
Background Characteristics		
Prior Preschool Experience		
Yes	11%	39%
No	89%	61%
Test Performance		
No differences found.		

APPENDIX F

PDC Classroom Observation System:
Definitions of Categories and
Means and Standard Deviations of Variables

Definitions of Observation Categories[1]

Category 1. Noninvolved:

The child is neither interacting with a person or object, nor engaged in a purposeful activity.

Category 2. Involved:

The child is interacting with a person or object. This category is also coded when the child is engaged in a purposeful activity such as singing to him/herself.

Subcategory 2a. Focus of Attention:

Social: The child is engaged in a reciprocal interaction with a peer and/or adult.

Nonsocial: The child is engaged in a purposeful activity which does not involve other persons.

Subcategory 2b. Language Spoken during Activity:

Verbal English: While engaging in social or nonsocial activities, the child speaks only in English.

Verbal Spanish: While engaging in social or nonsocial activities, the child speaks in English.

Verbal Combined: While engaging in social or nonsocial activities, the child uses a combination of Spanish and English words.

Nonverbal: While engaging in social or nonsocial activities, the child does not speak.

---

[1]For expanded definitions and examples of these categories, see Appendix E, Interim Report IV, Part A (August 1, 1976).

Category 3. Interactions with Peer:

The child engages in a reciprocal interaction with a peer(s) by looking at, listening to, talking with, or sharing materials and working on a common project.

Subcategory 3a. Description of Peer Interaction:

Negative: The child expresses verbal and/or physical aggression or hostility toward the peer(s).

Controlling: The child attempts, verbally or nonverbally, to direct, influence, or manipulate the behavior of a peer(s) in a positive manner. The child's behavior is directed toward one (or more) of the following outcomes: changing the peer's course of action, initiating a new behavior, or showing (telling) the peer how to do something.

Asserting: The child does not comply with or ignores in a positive manner attempts made by a peer to control his/her behavior.

Other: The child interacts in a cooperative and positive manner with a peer(s). The child is sharing, helping, taking turns, working jointly, listening to, or talking with a peer. (This item is only marked for each positive behavior that is clearly not a controlling and asserting behavior.)

Subcategories 3b and c. Purpose of Peer Interaction:

Requests Information: The child seeks (by posing a question, making a demand, etc.) a factual statement or explanation concerning a task, problem, casual relationship, or other events or situations in his/her environment.

Provides Information:	The child offers information in the form of factual statements, explanations, or physical gestures.
Requests Assistance:	The child seeks physical assistance or materials for initiating or completing an activity.
Provides Assistance:	The child offers physical assistance or materials for initiating or completing an activity.
Requests Support:	The child seeks comfort, protection, or reassurance after a hurt, disappointment, or other problem situation. The child does not request assistance or information for solving the problem.
Provides Support:	The child provides a peer with comfort, protection, or reassurance after a hurt, disappointment, or other problem situation. The child does not provide assistance or information for solving the problem. This item is also coded for verbal expressions of sympathy or empathy.
Nonapplicable:	The purpose of the child's interaction with a peer is clearly not one of requesting/providing information, assistance, materials, or emotional support.

Category 4. Interactions with Adults:

The child engages in a reciprocal interaction with an adult(s) by looking at, listening to, talking with, or sharing materials and working on a common project.

Subcategory 4a. Type of Adult Interaction:

Negative: The child expresses verbal and/or physical aggression or hostility toward the adult(s).

Controlling: The child attempts, verbally or nonverbally, to direct, influence, or manipulate the behavior of an adult(s) in a positive manner. The child's behavior is directed toward one (or more) of the following outcomes: changing the adult's course of action, initiating a new behavior, or showing (telling) the adult how to do something.

Asserting: The child does not comply with or ignores in a positive manner attempts made by an adult to control his/her behavior.

Other: The child interacts in a cooperative and positive manner with an adult(s). The child is sharing, helping, taking turns, working jointly, listening to, or talking with an adult. (This item is only marked for each positive behavior that is clearly not a controlling and asserting behavior).

Subcategories 4b and c. Purpose of Adult Interaction:

Requests Assistance: The child seeks physical assistance or materials for initiating or completing an activity.

Provides Assistance: The child offers physical assistance or materials for initiating or completing an activity.

Requests Support: The child seeks comfort, protection, or reassurance after a hurt, disappointment, or other problem situation. The child does not request assistance or information for solving the problem.

Provides Support: The child provides an adult with comfort, protection, or reassurance after a hurt, disappointment, or other problem situation. The child does not provide assistance or information for solving the problem. This item is also coded for verbal expressions of sympathy or empathy.

Nonapplicable: The purpose of the child's interaction with an adult is clearly not one of requesting/providing information, assistance, materials, or emotional support.

Category 5. Classroom Activity Level:

This category describes the opportunity children have to spontaneously engage in or initiate interactions with others. The observer's attention is no longer directed toward the target child, but toward the classroom as a whole. After surveying the classroom, one of the following items is coded after each 5-second observation interval:

Maximal: This refers to those parts of the day in which children and adults are free to initiate or maintain spontaneous interactions (verbally or physically) among themselves. The children are generally able to choose their own activity, with minimal structuring or direction by an adult. Teachers sometimes call these "free play" or "free choice" periods.

Moderate: This refers to those parts of the day in which the opportunity for spontaneous interaction among adults and children is substantially reduced. During this period, classroom behavior is typically decided less by children than by adult direction. There is still some opportunity for spontaneous interactions to occur within this structure.

Minimal: This refers to those parts of the day in which children are not free to initiate/maintain spontaneous interactions (verbal or physical) among themselves. Classroom behavior of the children is primarily controlled and directed by an adult.

Table F-1

Means and Standard Deviations of Classroom Child-Adult Interactions[a]

	ACROSS ALL activity levels		MAXIMUM activity levels		MODERATE activity levels		MINIMUM activity levels	
Observation Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Description of Child-Adult Interactions								
Negative	()[b]	()	()	()	()	()	()	()
Positive Control	.34	.20	.40	.30	.36	.28	.21	.26
Positive Assert	.02	.05	.02	.06	.02	.06	.02	.06
Positive Other	.65	.20	.58	.30	.63	.27	.77	.26
Purpose of Child-Adult Interactions								
Request Information	.08	.07	.11	.15	.08	.11	.04	.07
Give Information	.55	.19	.51	.26	.54	.27	.59	.35
Request Assistance/Materials	.13	.11	.18	.20	.14	.19	.08	.19
Give Assistance/Materials	.07	.12	.07	.12	.06	.14	.13	.26
Request Support	()	()	()	()	()	()	()	()
Give Support	()	()	()	()	()	()	()	()

[a]Estimates represent relative frequencies.

[b]Relative frequency of this category fell between .00 and .01.

Table F-2

Means and Standard Deviations of Classroom Involvement and Classroom Verbal Behavior Variables[a]

	ACROSS ALL activity levels		MAXIMUM activity levels		MODERATE activity levels		MINIMUM activity levels	
Observation Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Classroom Involvement								
Noninvolved	.06	.07	.05	.10	.07	.08	.08	.11
Social	.46	.16	.44	.19	.44	.19	.64	.24
Nonsocial	.48	.16	.51	.19	.49	.18	.28	.23
Classroom Verbal Behavior								
Verbal English	.27	.08	.29	.13	.25	.10	.19	.14
Verbal Spanish	.03	.08	.03	.08	.03	.09	.03	.07
Verbal Combination	()[b]	()	()	()	()	()	()	()
Nonverbal	.70	.08	.69	.13	.72	.11	.77	.15

[a]Estimates represent relative frequencies.

[b]Relative frequency of this category fell between .00 and .01.

Table F-3

Means and Standard Deviations of Classroom Child-Peer Interactions[a]

	ACROSS ALL activity levels		MAXIMUM activity levels		MODERATE activity levels		MINIMUM activity levels	
Observation Variable	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
Description of Child-Peer Interactions								
Negative	()[b]	()	.07	.01	()	()	.03	.12
Positive Control	.38	.22	.40	.24	.37	.29	.36	.36
Positive Assert	.03	.04	.04	.08	.02	.03	.04	.12
Positive Other	.58	.22	.55	.25	.61	.29	.57	.36
Purpose of Child-Peer Interactions								
Request Information	.08	.07	.08	.01	.11	.16	.05	.11
Give Information	.48	.18	.43	.20	.51	.23	.48	.35
Request Assistance/Materials	.17	.11	.20	.15	.14	.15	.20	.30
Give Assistance/Materials	.10	.12	.11	.16	.10	.15	.09	.17
Request Support	()	()	()	()	()	()	()	()
Give Support	()	()	()	()	()	()	()	()

[a]Estimates represent relative frequencies.

[b]Relative frequency of this category fell between .00 and .01.

APPENDIX G

Subscales of the POCL

Subscales of the POCL

Factor analysis[1] of the Pupil Observation Checklist yielded two distinct factors (accounting for 47.6% and 31.4% of the variance, respectively) whose composition is essentially identical to that of the factors discovered in past analyses of the POCL. As before, they have been named "Task Orientation" and "Extroversion" to describe the common characteristics of their constituent items. The loadings of each item on each factor are shown in Table G-1. Scores on the subscales were calculated by summing the actual scores on all items for a subscale. These subscale scores were then used for subsequent analyses of the POCL fall 1976 data.

Psychometric analyses of spring 1976 and fall 1976 POCL data indicate that the two subscales possess a high degree of internal consistency ($\alpha > .90$ at both time-points). Since this instrument is used by testers at all sites, for all groups, scores on these scales could be considered equivalent for English-dominant and Spanish-dominant samples, and might enter usefully into future analyses of data pooled across language groups.

---

[1] Principal components solution, varimax rotation.

Table G-1

Subscales of the POCL, Based on Factor Analysis[a] of Item Scores for All Children in the Analytic Sample Fall 1976 Data

N=916

POCL Item	Loading of Item on Each Factor (higher loading italicized)	
	Factor 1 "Task Orientation"	Factor 2 "Extroversion"
Cooperative	*.76*	.40
Sociable	.42	*.83*
Involved	*.69*	.55
Talkative	.32	*.87*
Attentive	*.83*	.26
Active	.25	*.86*
Relaxed	*.76*	.39
Quick to respond	*.62*	.52
Attempts difficult tasks	*.81*	.41
Keeps trying	*.86*	.29
Realistically self-confident	*.89*	.28

[a]Principal components solution, varimax rotation.

APPENDIX H

Analytic Procedure for Investigation of PDC and Comparison Classroom Comparability on the Observation Variables

Analytic Procedure for Investigation of PDC and Comparison Classroom Comparability on the Observation Variables

When analyzing data that represent proportions (such as the relative frequencies in the PDC Observation System), the data are sometimes transformed mathematically (e.g., arcsin of the square root of the proportion) to "normalize" the distributions of the proportion. These transformations are based on the assumption that the characteristic in question is distributed normally in the larger target population and thus should be forced into a normal configuration for the sample. For various reasons, that assumption is not tenable for these data. Consequently, analyses of PDC and comparison classroom comparability were performed using untransformed relative frequencies of the observation variables.

Since the purpose of performing these comparability analyses was to test the assumption that PDC and comparison group classroom means would not be significantly different, a liberal significance level of .10 was selected as the criterion for evaluating differences. Two group variables were created for these comparisons. In the first, all classrooms containing any children tested in English were split into PDC and comparison groups. In the second, all classrooms with any children tested in Spanish were divided into PDC and comparison groups.

An initial set of analyses was performed on the relative frequencies using the parametric t test. This statistic is robust with respect to violations of normality assumptions. However, in the event that the difference between group variances is significant, the results of such analyses must be interpreted cautiously. Consequently, a second set of non-parametric analyses, using the median test, were also performed for the same English-dominant sample (PDC vs. comparison) and Spanish-dominant sample (PDC vs. comparison).